

Attorney Docket No.: 015358-006600US  
Client Reference No.: ID-RSV-262

**NON-PROVISIONAL**

**PATENT APPLICATION**

**MULTIMEDIA VISUALIZATION AND INTEGRATION  
ENVIRONMENT**

Inventor: Jamey Graham, a citizen of The United States, residing at  
1196 Shasta Avenue  
San Jose, CA 95126

Assignee: Ricoh Company, Ltd.  
15-5 Minami Aoyama 1-Chome  
Minato-ku  
Tokyo, JAPAN 107-8544

Entity: Large

201220.62113001

## MULTIMEDIA VISUALIZATION AND INTEGRATION ENVIRONMENT

### COPYRIGHT

[01] A portion of the disclosure of this patent document contains material  
5 that is subject to copyright protection. The copyright owner has no objection to the  
xerographic reproduction by anyone of the patent document or the patent disclosure in  
exactly the form it appears in the U.S. Patent and Trademark Office patent file or records, but  
otherwise reserves all copyright rights whatsoever.

### CROSS-REFERENCES TO RELATED APPLICATIONS

[02] The present application claims priority from and is a continuation-in-  
part (CIP) application of U.S. Non-Provisional Patent Application No. 08/995,616, entitled  
“AUTOMATIC ADAPTIVE DOCUMENT READING HELP SYSTEM” filed December  
22, 1997, the entire contents of which are herein incorporated by reference for all purposes.

[03] The present application incorporates by reference for all purposes the  
entire contents of U.S. Non-Provisional Application No. 10/001,895 (Attorney Docket No.:  
15358-006500US), entitled “PAPER-BASED INTERFACE FOR MULTIMEDIA  
INFORMATION” filed November 19, 2001.

### BACKGROUND OF THE INVENTION

[04] The present invention relates to user interfaces for displaying  
information and more particularly to user interfaces for retrieving and displaying multimedia  
information that may be stored in a multimedia document.

[05] With rapid advances in computer technology, an increasing amount of  
25 information is being stored in the form of electronic (or digital) documents. These electronic  
documents include multimedia documents that store multimedia information. The term  
“multimedia information” is used to refer to information that comprises information of  
several different types in an integrated form. The different types of information included in  
multimedia information may include a combination of text information, graphics information,  
30 animation information, sound (audio) information, video information, slides information,  
whiteboard information, and other types of information. Multimedia information is also used  
to refer to information comprising one or more objects wherein the objects include

information of different types. For example, multimedia objects included in multimedia information may comprise text information, graphics information, animation information, sound (audio) information, video information, slides information, whiteboard information, and other types of information. Multimedia documents may be considered as compound objects that comprise video, audio, closed-caption text, keyframes, presentation slides, whiteboard capture information, as well as other multimedia type objects. Examples of multimedia documents include documents storing interactive web pages, television broadcasts, videos, presentations, or the like.

[06] Several tools and applications are conventionally available that allow users to play back, store, index, edit, or manipulate multimedia information stored in multimedia documents. Examples of such tools and/or applications include proprietary or customized multimedia players (e.g., RealPlayer™ provided by RealNetworks, Microsoft Windows Media Player provided by Microsoft Corporation, QuickTime™ Player provided by Apple Corporation, Shockwave multimedia player, and others), video players, televisions, personal digital assistants (PDAs), or the like. Several tools are also available for editing multimedia information. For example, Virage, Inc. of San Mateo, California (www.virage.com) provides various tools for viewing and manipulating video content and tools for creating video databases. Virage, Inc. also provides tools for face detection and on-screen text recognition from video information.

[07] Given the vast number of electronic documents, readers of electronic documents are increasingly being called upon to assimilate vast quantities of information in a short period of time. To meet the demands placed upon them, readers find they must read electronic documents “horizontally” rather than “vertically,” i.e., they must scan, skim, and browse sections of interest in one or more electronic documents rather than read and analyze a single document from start to end. While tools exist which enable users to “horizontally” read electronic documents containing text/image information (e.g., the reading tool described in U.S. Non-Provisional Patent Application No. 08/995,616), conventional tools cannot be used to “horizontally” read multimedia documents which may contain audio information, video information, and other types of information. None of the multimedia tools described above allow users to “horizontally” read a multimedia document.

In light of the above, there is a need for techniques that allow users to read a multimedia document “horizontally.” Techniques that allow users to view, analyze, and navigate multimedia information stored in multimedia documents are desirable.

## BRIEF SUMMARY OF THE INVENTION

[08] The present invention provides techniques for retrieving and displaying multimedia information. According to an embodiment of the present invention, a graphical user interface (GUI) is provided that displays multimedia information that may be stored in a multimedia document. According to the teachings of the present invention, the GUI enables a user to navigate through multimedia information stored in a multimedia document. The GUI provides both a focused and a contextual view of the contents of the multimedia document. The GUI thus allows users to “horizontally” read multimedia documents.

[09] According to an embodiment of the present invention, techniques are provided for displaying multimedia information stored in a multimedia document on a display. The multimedia information comprises information of a plurality of types including information of a first type and information of a second type. In this embodiment, a graphical user interface (GUI) is displayed on the display. A representation of the multimedia information stored by the multimedia document is displayed in a first area of the GUI. The displayed representation of the multimedia information in the first area comprises a representation of information of the first type and a representation of information of the second type. A first lens is displayed covering a first portion of the first area. A representation of multimedia information comprising a portion of the representation of information of the first type covered by the first lens and a portion of the representation of information of the second type covered by the first lens is displayed in a second area of the GUI.

[10] According to another embodiment of the present invention, techniques are provided for displaying multimedia information stored in a multimedia document on a display. The multimedia information comprises information of a first type and information of a second type. In this embodiment, a graphical user interface (GUI) is displayed on the display. A representation of the multimedia information stored by the multimedia document occurring between a start time ( $t_s$ ) and an end time ( $t_e$ ) associated with the multimedia document is displayed in a first area of the GUI. The representation of the multimedia information displayed in the first area of the GUI comprises a representation of information of the first type occurring between  $t_s$  and  $t_e$  and a representation of information of the second type occurring between  $t_s$  and  $t_e$ , where ( $t_e > t_s$ ). A first lens is displayed emphasizing a portion of the first area of the GUI, where the portion of the first area emphasized by the first lens comprises a representation of multimedia information occurring between a first time ( $t_1$ )

and a second time ( $t_2$ ), where ( $t_s \leq t_1 < t_2 \leq t_e$ ). The representation of multimedia information occurring between  $t_1$  and  $t_2$  is displayed in a second area of the GUI. The representation of multimedia information displayed in the second area comprises a representation of information of the first type occurring between  $t_1$  and  $t_2$  and a representation of information of the second type occurring between  $t_1$  and  $t_2$ .

[11] According to yet another embodiment of the present invention, techniques are provided for displaying multimedia information stored in a multimedia document on a display. The multimedia information comprises video information and information of a first type. In this embodiment, a graphical user interface (GUI) is displayed on the display. A first set of one or more video keyframes extracted from the video information occurring between a start time ( $t_s$ ) and an end time ( $t_e$ ) associated with the multimedia document, where ( $t_e > t_s$ ), are displayed in a first section of a first area of the GUI. Text information corresponding to the information of the first type occurring between  $t_s$  and  $t_e$  is displayed in a second section of the first area of the GUI. A first lens is displayed emphasizing a portion of the first section of the first area occurring between a first time ( $t_1$ ) and a second time ( $t_2$ ) and a portion of the second section of the first area occurring between  $t_1$  and  $t_2$ . The emphasized portion of the first section of the first area comprises a second set of one or more video keyframes extracted from the video information occurring between  $t_1$  and  $t_2$ , and the emphasized portion of the second section of the first area comprises text information corresponding to information of the first type occurring between  $t_1$  and  $t_2$ , wherein the second set of one or more keyframes is a subset of the first set of one or more keyframes and ( $t_s \leq t_1 < t_2 \leq t_e$ ). The second set of one or more keyframes is displayed in a first section of a second area of the GUI. Text information corresponding to the information of the first type occurring between  $t_1$  and  $t_2$  is displayed in a second section of the second area of the GUI.

[12] The foregoing, together with other features, embodiments, and advantages of the present invention, will become more apparent when referring to the following specification, claims, and accompanying drawings.

## BRIEF DESCRIPTION OF THE DRAWINGS

[13] Fig. 1 is a simplified block diagram of a distributed network that may incorporate an embodiment of the present invention;

[14] Fig. 2 is a simplified block diagram of a computer system according to an embodiment of the present invention;

[15] Fig. 3 depicts a simplified user interface 300 generated according to an embodiment of the present invention for viewing multimedia information;

[16] Fig. 4 is a zoomed-in simplified diagram of a thumbnail viewing area lens according to an embodiment of the present invention;

[17] Figs. 5A, 5B, and 5C are simplified diagrams of a panel viewing area lens according to an embodiment of the present invention;

[18] Fig. 6 depicts a simplified user interface generated according to an embodiment of the present invention wherein user-selected words are annotated;

[19] Fig. 7 is a simplified zoomed-in view of a second viewing area of a GUI generated according to an embodiment of the present invention;

[20] Fig. 8 depicts a simplified GUI in which multimedia information that is relevant to one or more topics of interest to a user is annotated according to an embodiment of the present invention;

[21] Fig. 9 depicts a simplified user interface for defining a topic of interest according to an embodiment of the present invention;

[22] Fig. 10 depicts a simplified user interface that displays multimedia information stored by a meeting recording according to an embodiment of the present invention;

[23] Fig. 11 depicts a simplified user interface that displays multimedia information stored by a multimedia document according to an embodiment of the present invention;

[24] Fig. 12 depicts a simplified user interface that displays multimedia information stored by a multimedia document according to an embodiment of the present invention;

[25] Fig. 13 depicts a simplified user interface that displays contents of a multimedia document according to an embodiment of the present invention;

[26] Fig. 14 is a simplified high-level flowchart depicting a method of displaying a thumbnail depicting text information in the second viewing area of a GUI according to an embodiment of the present invention;

10 [27] Fig. 15 is a simplified high-level flowchart depicting a method of displaying a thumbnail that depicts video keyframes extracted from the video information in the second viewing area of a GUI according to an embodiment of the present invention;

5 [28] Fig. 16 is a simplified high-level flowchart depicting another method of displaying thumbnail 312-2 according to an embodiment of the present invention;

10 [29] Fig. 17 is a simplified high-level flowchart depicting a method of displaying thumbnail viewing area lens 314, displaying information emphasized by thumbnail viewing area lens 314 in third viewing area 306, displaying panel viewing area lens 322, displaying information emphasized by panel viewing area lens 322 in fourth viewing area 308, and displaying information in fifth viewing area 310 according to an embodiment of the present invention;

15 [30] Fig. 18 is a simplified high-level flowchart depicting a method of automatically updating the information displayed in third viewing area 306 in response to a change in the location of thumbnail viewing area lens 314 according to an embodiment of the present invention;

20 [31] Fig. 19 is a simplified high-level flowchart depicting a method of automatically updating the information displayed in fourth viewing area 308 and the positions of thumbnail viewing area lens 314 and sub-lens 316 in response to a change in the location of panel viewing area lens 322 according to an embodiment of the present invention;

25 [32] Fig. 20A depicts a simplified user interface that displays ranges according to an embodiment of the present invention;

[33] Fig. 20B depicts a simplified dialog box for editing ranges according to an embodiment of the present invention;

30 [34] Fig. 21 is a simplified high-level flowchart depicting a method of automatically creating ranges according to an embodiment of the present invention;

[35] Fig. 22 is a simplified high-level flowchart depicting a method of automatically creating ranges based upon locations of hits in the multimedia information according to an embodiment of the present invention;

35 [36] Fig. 23 is a simplified high-level flowchart depicting a method of combining one or more ranges based upon the size of the ranges and the proximity of the ranges to neighboring ranges according to an embodiment of the present invention;

[37] Fig. 24 depicts a simplified diagram showing the relationships between neighboring ranges according to an embodiment of the present invention;

[38] Fig. 25A depicts a simplified diagram showing a range created by combining ranges  $R_i$  and  $R_k$  depicted in Fig. 24 according to an embodiment of the present invention;

[39] Fig. 25B depicts a simplified diagram showing a range created by combining ranges  $R_i$  and  $R_j$  depicted in Fig. 24 according to an embodiment of the present invention; and

[40] Fig. 26 depicts a zoomed-in version of a GUI depicting ranges that have been automatically created according to an embodiment of the present invention.

#### DETAILED DESCRIPTION OF THE INVENTION

[41] Embodiments of the present invention provide techniques for retrieving and displaying multimedia information. According to an embodiment of the present invention, a graphical user interface (GUI) is provided that displays multimedia information that may be stored in a multimedia document. According to the teachings of the present invention, the GUI enables a user to navigate through multimedia information stored in a multimedia document. The GUI provides both a focused and a contextual view of the contents of the multimedia document. The GUI thus allows a user to “horizontally” read multimedia documents.

[42] As indicated above, the term “multimedia information” is intended to refer to information that comprises information of several different types in an integrated form. The different types of information included in multimedia information may include a combination of text information, graphics information, animation information, sound (audio) information, video information, slides information, whiteboard images information, and other types of information. For example, a video recording of a television broadcast may comprise video information and audio information. In certain instances the video recording may also comprise close-captioned (CC) text information which comprises material related to the video information, and in many cases, is an exact representation of the speech contained in the audio portions of the video recording. Multimedia information is also used to refer to information comprising one or more objects wherein the objects include information of different types. For example, multimedia objects included in multimedia information may comprise text information, graphics information, animation information, sound (audio) information, video information, slides information, whiteboard images information, and other types of information.



[43] The term "multimedia document" as used in this application is intended to refer to any electronic storage unit (e.g., a file) that stores multimedia information in digital format. Various different formats may be used to store the multimedia information. These formats include various MPEG formats (e.g., MPEG 1, MPEG 2, MPEG 4, MPEG 7, etc.), MP3 format, SMIL format, HTML+TIME format, WMF (Windows Media Format), RM (Real Media) format, Quicktime format, Shockwave format, various streaming media formats, formats being developed by the engineering community, proprietary and customary formats, and others. Examples of multimedia documents include video recordings, MPEG files, news broadcast recordings, presentation recordings, recorded meetings, classroom lecture recordings, broadcast television programs, or the like.

[44] Fig. 1 is a simplified block diagram of a distributed network 100 that may incorporate an embodiment of the present invention. As depicted in Fig. 1, distributed network 100 comprises a number of computer systems including one or more client systems 102, a server system 104, and a multimedia information source (MIS) 106 coupled to communication network 108 via a plurality of communication links 110. Distributed network 100 depicted in Fig. 1 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. For example, the present invention may also be embodied in a stand-alone system. In a stand-alone environment, the functions performed by the various computer systems depicted in Fig. 1 may be performed by a single computer system.

[45] Communication network 108 provides a mechanism allowing the various computer systems depicted in Fig. 1 to communicate and exchange information with each other. Communication network 108 may itself be comprised of many interconnected computer systems and communication links. While in one embodiment, communication network 108 is the Internet, in other embodiments, communication network 108 may be any suitable communication network including a local area network (LAN), a wide area network (WAN), a wireless network, an intranet, a private network, a public network, a switched network, or the like.

[46] Communication links 110 used to connect the various systems depicted in Fig. 1 may be of various types including hardwire links, optical links, satellite or other wireless communications links, wave propagation links, or any other mechanisms for communication of information. Various communication protocols may be used to facilitate communication of information via the communication links. These communication protocols

may include TCP/IP, HTTP protocols, extensible markup language (XML), wireless application protocol (WAP), protocols under development by industry standard organizations, vendor-specific protocols, customized protocols, and others.

[47] Computer systems connected to communication network 108 may be classified as "clients" or "servers" depending on the role the computer systems play with respect to requesting information and/or services or providing information and/or services. Computer systems that are used by users to request information or to request a service are classified as "client" computers (or "clients"). Computer systems that store information and provide the information in response to a user request received from a client computer, or computer systems that perform processing to provide the user-requested services are called "server" computers (or "servers"). It should however be apparent that a particular computer system may function both as a client and as a server.

[48] Accordingly, according to an embodiment of the present invention, server system 104 is configured to perform processing to facilitate generation of a GUI that displays multimedia information according to the teachings of the present invention. The GUI generated by server system 104 may be output to the user (e.g., a reader of the multimedia document) via an output device coupled to server system 104 or via client systems 102. The GUI generated by server 104 enables the user to retrieve and browse multimedia information that may be stored in a multimedia document. The GUI provides both a focused and a contextual view of the contents of a multimedia document and thus enables the multimedia document to be read "horizontally."

[49] The processing performed by server system 104 to generate the GUI and to provide the various features according to the teachings of the present invention may be implemented by software modules executing on server system 104, by hardware modules coupled to server system 104, or combinations thereof. In alternative embodiments of the present invention, the processing may also be distributed between the various computer systems depicted in Fig. 1.

[50] The multimedia information that is displayed in the GUI may be stored in a multimedia document that is accessible to server system 104. For example, the multimedia document may be stored in a storage subsystem of server system 104. The multimedia document may also be stored by other systems such as MIS 106 that are accessible to server 104. Alternatively, the multimedia document may be stored in a memory location accessible to server system 104.

[51] In alternative embodiments, instead of accessing a multimedia document, server system 104 may receive a stream of multimedia information (e.g., a streaming media signal, a cable signal, etc.) from a multimedia information source such as MIS 106. According to an embodiment of the present invention, server system 104 stores the multimedia information signals in a multimedia document and then generates a GUI that displays the multimedia information. Examples of MIS 106 include a television broadcast receiver, a cable receiver, a digital video recorder (e.g., a TIVO box), or the like. For example, multimedia information source 106 may be embodied as a television that is configured to receive multimedia broadcast signals and to transmit the signals to server system 104. In alternative embodiments, server system 104 may be configured to intercept multimedia information signals received by MIS 106. Server system 104 may receive the multimedia information directly from MIS 106 or may alternatively receive the information via a communication network such as communication network 108.

[52] As described above, MIS 106 depicted in Fig. 1 represents a source of multimedia information. According to an embodiment of the present invention, MIS 106 may store multimedia documents that are accessed by server system 104. For example, MIS 106 may be a storage device or a server that stores multimedia documents that may be accessed by server system 104. In alternative embodiments, MIS 106 may provide a multimedia information stream to server system 104. For example, MIS 106 may be a television receiver/antenna providing live television feed information to server system 104. MIS 106 may be a device such as a video recorder/player, a DVD player, a CD player, etc. providing recorded video and/or audio stream to server system 104. In alternative embodiments, MIS 106 may be a presentation or meeting recorder device that is capable of providing a stream of the captured presentation or meeting information to server system 104. MIS 106 may also be a receiver (e.g., a satellite dish or a cable receiver) that is configured to capture or receive (e.g., via a wireless link) multimedia information from an external source and then provide the captured multimedia information to server system 104 for further processing.

[53] Users may use client systems 102 to view the GUI generated by server system 104. Users may also use client systems 102 to interact with the other systems depicted in Fig. 1. For example, a user may use user system 102 to select a particular multimedia document and request server system 104 to generate a GUI displaying multimedia information stored by the particular multimedia document. A user may also interact with the GUI generated by server system 104 using input devices coupled to client

system 102. In alternative embodiments, client system 102 may also perform processing to facilitate generation of a GUI according to the teachings of the present invention. A client system 102 may be of different types including a personal computer, a portable computer, a workstation, a computer terminal, a network computer, a mainframe, a kiosk, a personal digital assistant (PDA), a communication device such as a cell phone, or any other data processing system.

[54] According to an embodiment of the present invention, a single computer system may function both as server system 104 and as client system 102. Various other configurations of the server system 104, client system 102, and MIS 106 are possible.

[55] Fig. 2 is a simplified block diagram of a computer system 200 according to an embodiment of the present invention. Computer system 200 may be used as any of the computer systems depicted in Fig. 1. As shown in Fig. 2, computer system 200 includes at least one processor 202, which communicates with a number of peripheral devices via a bus subsystem 204. These peripheral devices may include a storage subsystem 206, comprising a memory subsystem 208 and a file storage subsystem 210, user interface input devices 212, user interface output devices 214, and a network interface subsystem 216. The input and output devices allow user interaction with computer system 200. A user may be a human user, a device, a process, another computer, or the like. Network interface subsystem 216 provides an interface to other computer systems and communication networks.

[56] Bus subsystem 204 provides a mechanism for letting the various components and subsystems of computer system 200 communicate with each other as intended. The various subsystems and components of computer system 200 need not be at the same physical location but may be distributed at various locations within network 100. Although bus subsystem 204 is shown schematically as a single bus, alternative embodiments of the bus subsystem may utilize multiple busses.

[57] User interface input devices 212 may include a keyboard, pointing devices, a mouse, trackball, touchpad, a graphics tablet, a scanner, a barcode scanner, a touchscreen incorporated into the display, audio input devices such as voice recognition systems, microphones, and other types of input devices. In general, use of the term "input device" is intended to include all possible types of devices and ways to input information using computer system 200.

[58] User interface output devices 214 may include a display subsystem, a printer, a fax machine, or non-visual displays such as audio output devices. The display subsystem may be a cathode ray tube (CRT), a flat-panel device such as a liquid crystal

display (LCD), a projection device, or the like. The display subsystem may also provide non-visual display such as via audio output devices. In general, use of the term "output device" is intended to include all possible types of devices and ways to output information from computer system 200. According to an embodiment of the present invention, the GUI generated according to the teachings of the present invention may be presented to the user via output devices 214.

[59] Storage subsystem 206 may be configured to store the basic programming and data constructs that provide the functionality of the computer system and of the present invention. For example, according to an embodiment of the present invention, software modules implementing the functionality of the present invention may be stored in storage subsystem 206 of server system 104. These software modules may be executed by processor(s) 202 of server system 104. In a distributed environment, the software modules may be stored on a plurality of computer systems and executed by processors of the plurality of computer systems. Storage subsystem 206 may also provide a repository for storing various databases that may be used by the present invention. Storage subsystem 206 may comprise memory subsystem 208 and file storage subsystem 210.

[60] Memory subsystem 208 may include a number of memories including a main random access memory (RAM) 218 for storage of instructions and data during program execution and a read only memory (ROM) 220 in which fixed instructions are stored. File storage subsystem 210 provides persistent (non-volatile) storage for program and data files, and may include a hard disk drive, a floppy disk drive along with associated removable media, a Compact Disk Read Only Memory (CD-ROM) drive, an optical drive, removable media cartridges, and other like storage media. One or more of the drives may be located at remote locations on other connected computers.

[61] Computer system 200 itself can be of varying types including a personal computer, a portable computer, a workstation, a computer terminal, a network computer, a mainframe, a kiosk, a personal digital assistant (PDA), a communication device such as a cell phone, or any other data processing system. Server computers generally have more storage and processing capacity than client systems. Due to the ever-changing nature of computers and networks, the description of computer system 200 depicted in Fig. 2 is intended only as a specific example for purposes of illustrating the preferred embodiment of the computer system. Many other configurations of a computer system are possible having more or fewer components than the computer system depicted in Fig. 2.

[62] Fig. 3 depicts a simplified user interface 300 generated according to an embodiment of the present invention for viewing multimedia information. It should be apparent that GUI 300 depicted in Fig. 3 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[63] GUI 300 displays multimedia information stored in a multimedia document. The multimedia information stored by the multimedia document and displayed by GUI 300 may comprise information of a plurality of different types. As depicted in Fig. 3, GUI 300 displays multimedia information corresponding to a television broadcast that includes video information, audio information, and possibly closed-caption (CC) text information. The television broadcast may be stored as a television broadcast recording in a memory location accessible to server system 104. It should however be apparent that the present invention is not restricted to displaying television recordings. Multimedia information comprising other types of information may also be displayed according to the teachings of the present invention.

[64] The television broadcast may be stored using a variety of different techniques. According to one technique, the television broadcast is recorded and stored using a satellite receiver connected to a PC-TV video card of server system 104. Applications executing on server system 104 then process the recorded television broadcast to facilitate generation of GUI 300. For example, the video information contained in the television broadcast may be captured using an MPEG capture application that creates a separate metafile (e.g., in XML format) containing temporal information for the broadcast and closed-caption text, if provided. Information stored in the metafile may then be used to generate GUI 300 depicted in Fig. 3.

[65] As depicted in Fig. 3, GUI 300 comprises several viewing areas including a first viewing area 302, a second viewing area 304, a third viewing area 306, a fourth viewing area 308, and a fifth viewing area 310. It should be apparent that in alternative embodiments the present invention may comprise more or fewer viewing areas than those depicted in Fig. 3. Further, in alternative embodiments of the present invention one or more viewing areas may be combined into one viewing area, or a particular viewing area may be divided in multiple viewing areas. Accordingly, the viewing areas depicted in Fig. 3 and described below are not meant to restrict the scope of the present invention as recited in the claims.

10 [66] According to an embodiment of the present invention, first viewing area 302 displays one or more commands that may be selected by a user viewing GUI 300. Various user interface features such as menu bars, drop-down menus, cascading menus, buttons, selection bars, buttons, etc. may be used to display the user-selectable commands.

5 According to an embodiment of the present invention, the commands provided in first viewing area 302 include a command that enables the user to select a multimedia document whose multimedia information is to be displayed in the GUI. The commands may also include one or more commands that allow the user to configure and/or customize the manner in which multimedia information stored in the user-selected multimedia document is displayed in GUI 300. Various other commands may also be provided in first viewing area 302.

15 [67] According to an embodiment of the present invention, second viewing area 304 displays a scaled representation of multimedia information stored by the multimedia document. The user may select the scaling factor used for displaying information in second viewing area 304. According to a particular embodiment of the present invention, a representation of the entire (i.e., multimedia information between the start time and end time associated with the multimedia document) multimedia document is displayed in second viewing area 304. In this embodiment, one end of second viewing area 304 represents the start time of the multimedia document and the opposite end of second viewing area 304 represents the end time of the multimedia document.

20 [68] As shown in Fig. 3, according to an embodiment of the present invention, second viewing area 304 comprises one or more thumbnail images 312. Each thumbnail image displays a representation of a particular type of information included in the multimedia information stored by the multimedia document. For example, two thumbnail images 312-1 and 312-2 are displayed in second viewing area 304 of GUI 300 depicted in Fig. 3. Thumbnail image 312-1 displays text information corresponding to information included in the multimedia information stored by the multimedia document being displayed by GUI 300. The text displayed in thumbnail image 312-1 may represent a displayable representation of CC text included in the multimedia information displayed by GUI 300.

25 Alternatively, the text displayed in thumbnail image 312-1 may represent a displayable representation of a transcription of audio information included in the multimedia information stored by the multimedia document whose contents are displayed by GUI 300. Various audio-to-text transcription techniques may be used to generate a transcript for the audio information.

[69] Thumbnail image 312-2 displays a representation of video information included in the multimedia information displayed by GUI 300. In the embodiment depicted in Fig. 3, the video information is displayed using video keyframes extracted from the video information included in the multimedia information stored by the multimedia document. The video keyframes may be extracted from the video information in the multimedia document at various points in time using a specified sampling rate. A special layout style, which may be user-configurable, is used to display the extracted keyframes in thumbnail image 312-2 to enhance readability of the frames.

[70] One or more thumbnail images may be displayed in second viewing area 304 based upon the different types of information included in the multimedia information being displayed. Each thumbnail image 312 displayed in second viewing area 304 displays a representation of information of a particular type included in the multimedia information stored by the multimedia document. According to an embodiment of the present invention, the number of thumbnails displayed in second viewing area 304 and the type of information displayed by each thumbnail is user-configurable.

[71] According to an embodiment of the present invention, the various thumbnail images displayed in second viewing area 304 are temporally synchronized or aligned with each other along a timeline. This implies that the various types of information included in the multimedia information and occurring at approximately the same time are displayed next to each other. For example, thumbnail images 312-1 and 312-2 are aligned such that the text information (which may represent CC text information or a transcript of the audio information) displayed in thumbnail image 312-1 and video keyframes displayed in thumbnail 312-2 that occur in the multimedia information at a particular point in time are displayed close to each other (e.g., along the same horizontal axis). Accordingly, information that has a particular time stamp is displayed proximal to information that has approximately the same time stamp. This enables a user to determine the various types of information occurring approximately concurrently in the multimedia information being displayed by GUI 300 by simply scanning second viewing area 304 in the horizontal axis.

[72] According to the teachings of the present invention, a viewing lens or window 314 (hereinafter referred to as "thumbnail viewing area lens 314") is displayed in second viewing area 304. Thumbnail viewing area lens 314 covers or emphasizes a portion of second viewing area 304. According to the teachings of the present invention, multimedia information corresponding to the area of second viewing area 304 covered by thumbnail viewing area lens 314 is displayed in third viewing area 306.



[73] In the embodiment depicted in Fig. 3, thumbnail viewing area lens 314 is positioned at the top of second viewing area 304 and emphasizes a top portion (or starting portion) of the multimedia document. The position of thumbnail viewing area lens 314 may be changed by a user by sliding or moving lens 314 along second viewing area 304. For example, in Fig. 3, thumbnail viewing area lens 314 may be moved vertically along second viewing area 304.

[74] In response to a change in the position of thumbnail viewing area lens 314 from a first location in second viewing area 304 to a second location along second viewing area 304, the multimedia information displayed in third viewing area 306 is automatically updated such that the multimedia information displayed in third viewing area 306 continues to correspond to the area of second viewing area 304 emphasized by thumbnail viewing area lens 314. Accordingly, a user may use thumbnail viewing area lens 314 to navigate and scroll through the contents of the multimedia document displayed by GUI 300. Thumbnail viewing area lens 314 thus provides a context and indicates a location of the multimedia information displayed in third viewing area 306 within the entire multimedia document.

[75] Fig. 4 is a zoomed-in simplified diagram of thumbnail viewing area lens 314 according to an embodiment of the present invention. As depicted in Fig. 4, thumbnail viewing area lens 314 is bounded by a first edge 318 and a second edge 320. Thumbnail viewing area lens 314 emphasizes an area of second viewing area 304 between edge 318 and edge 320. Based upon the position of thumbnail viewing area lens 314 over second viewing area 304, edge 318 corresponds to specific time "t<sub>1</sub>" in the multimedia document and edge 320 corresponds to a specific time "t<sub>2</sub>" in the multimedia document wherein t<sub>2</sub> > t<sub>1</sub>. For example, when thumbnail viewing area lens 314 is positioned at the start of second viewing area 304 (as depicted in Fig. 3), t<sub>1</sub> may correspond to the start time of the multimedia document being displayed, and when thumbnail viewing area lens 314 is positioned at the end of second viewing area 304, t<sub>2</sub> may correspond to the end time of the multimedia document. Accordingly, thumbnail viewing area lens 314 emphasizes a portion of second viewing area 304 between times t<sub>1</sub> and t<sub>2</sub>. According to an embodiment of the present invention, multimedia information corresponding to the time segment between t<sub>2</sub> and t<sub>1</sub> (which is emphasized or covered by thumbnail viewing area lens 314) is displayed in third viewing area 306. Accordingly, when the position of thumbnail viewing area lens 314 is changed along second viewing area 304 in response to user input, the information displayed in third viewing area 306 is updated such that the multimedia information displayed in third

viewing area 306 continues to correspond to the area of second viewing area 304 emphasized by thumbnail viewing area lens 314.

[76] As shown in Fig. 4 and Fig. 3, thumbnail viewing area lens 314 comprises a sub-lens 316 which further emphasizes a sub-portion of the portion of second viewing area 304 emphasized by thumbnail viewing area lens 314. According to an embodiment the present invention, the portion of second viewing area 304 emphasized or covered by sub-lens 316 corresponds to the portion of third viewing area 306 emphasized by lens 322. Sub-lens 316 can be moved along second viewing area 304 within edges 318 and 320 of thumbnail viewing area lens 314. When sub-lens 316 is moved from a first location to a second location within the boundaries of thumbnail viewing area lens 314, the position of lens 322 in third viewing area 306 is also automatically changed to correspond to the changed location of sub-lens 316. Further, if the position of lens 322 is changed from a first location to a second location over third viewing area 306, the position of sub-lens 316 is also automatically updated to correspond to the changed position of lens 322. Further details related to lens 322 are described below.

[77] As described above, multimedia information corresponding to the portion of second viewing area 304 emphasized by thumbnail viewing area lens 314 is displayed in third viewing area 306. Accordingly, a representation of multimedia information occurring between time  $t_1$  and  $t_2$  (corresponding to a segment of time of the multimedia document emphasized by thumbnail viewing area lens 314) is displayed in third viewing area 306. Third viewing area 306 thus displays a zoomed-in representation of the multimedia information stored by the multimedia document corresponding to the portion of the multimedia document emphasized by thumbnail viewing area lens 314.

[78] As depicted in Fig. 3, third viewing area 306 comprises one or more panels 324. Each panel displays a representation of information of a particular type included in the multimedia information occurring during the time segment emphasized by thumbnail viewing area lens 314. For example, in GUI 300 depicted in Fig. 3, two panels 324-1 and 324-2 are displayed in third viewing area 306. According to an embodiment of the present invention, each panel 324 in third viewing area 306 corresponds to a thumbnail image 312 displayed in second viewing area 304 and displays information corresponding to the section of the thumbnail image covered by thumbnail viewing area lens 314.

[79] Like thumbnail images 312, panels 324 are also temporally aligned or synchronized with each other. Accordingly, the various types of information included in the multimedia information and occurring at approximately the same time are displayed next to

each other in third viewing area 306. For example, panels 324-1 and 324-2 depicted in Fig. 3 are aligned such that the text information (which may represent CC text information or a transcript of the audio information) displayed in panel 324-1 and video keyframes displayed in panel 324-2 that occur in the multimedia information at a approximately the same point in time are displayed close to each other (e.g., along the same horizontal axis). Accordingly, information that has a particular time stamp is displayed proximal to other types of information that has approximately the same time stamp. This enables a user to determine the various types of information occurring approximately concurrently in the multimedia information by simply scanning third viewing area 306 in the horizontal axis.

[80] Panel 324-1 depicted in GUI 300 corresponds to thumbnail image 312-1 and displays text information corresponding to the area of thumbnail image 312-1 emphasized or covered by thumbnail viewing area lens 314. The text information displayed by panel 324-1 may correspond to text extracted from CC information included in the multimedia information, or alternatively may represent a transcript of audio information included in the multimedia information. According to an embodiment of the present invention, the present invention takes advantage of the automatic story segmentation and other features that are often provided in close-captioned (CC) text from broadcast news. Most news agencies who provide CC text as part of their broadcast use a special syntax in the CC text (e.g., a ">>>>" delimiter to indicate changes in story line or subject, a ">>" delimiter to indicate changes in speakers, etc.). Given the presence of this kind of information in the CC text information included in the multimedia information, the present invention incorporates these features in the text displayed in panel 324-1. For example, a ">>>>" delimiter may be displayed to indicate changes in story line or subject, a ">>" delimiter may be displayed to indicate changes in speakers, additional spacing may be displayed between text portions related to different story lines to clearly demarcate the different stories, etc. This enhances the readability of the text information displayed in panel 324-1.

[81] Panel 324-2 depicted in GUI 300 corresponds to thumbnail image 312-2 and displays a representation of video information corresponding to the area of thumbnail image 312-2 emphasized or covered by thumbnail viewing area lens 314. Accordingly, panel 324-2 displays a representation of video information included in the multimedia information stored by the multimedia document and occurring between times  $t_1$  and  $t_2$  associated with thumbnail viewing area lens 314. In the embodiment depicted in Fig. 3, video keyframes extracted from the video information included in the multimedia information are displayed in

panel 324-2. A special layout style (which is user-configurable) is used to display the extracted keyframes to enhance readability of the frames.

[82] Various different techniques may be used to display video keyframes in panel 324-2. According to an embodiment of the present invention, the time segment between time  $t_1$  and time  $t_2$  is divided into sub-segments of a pre-determined time period. Each sub-segment is characterized by a start time and an end time associated with the sub-segment. According to an embodiment of the present invention, the start time of the first sub-segment corresponds to time  $t_1$  while the end time of the last sub-segment corresponds to time  $t_2$ . Server 104 then extracts a set of one or more video keyframes from the video information stored by the multimedia document for each sub-segment occurring between the start time and end time associated with the sub-segment. For example, according to an embodiment of the present invention, for each sub-segment, server 104 may extract a video keyframe at 1-second intervals between a start time and an end time associated with the sub-segment.

[83] For each sub-segment, server 104 then selects one or more keyframes from the set of extracted video keyframes for the sub-segment to be displayed in panel 324-2. The number of keyframes selected to be displayed in panel 324-2 for each sub-segment is user-configurable. Various different techniques may be used for selecting the video keyframes to be displayed from the extracted set of video keyframes for each time sub-segment. For example, if the set of video keyframes extracted for a sub-segment comprises 24 keyframes and if six video keyframes are to be displayed for each sub-segment (as shown in Fig. 3), server 104 may select the first two video keyframes, the middle two video keyframes, and the last two video keyframes from the set of extracted video keyframes for the sub-segment.

[84] In another embodiment, the video keyframes to be displayed for a sub-segment may be selected based upon the sequential positions of the keyframes in the set of keyframes extracted for sub-segment. For example, if the set of video keyframes extracted for a sub-segment comprises 24 keyframes and if six video keyframes are to be displayed for each sub-segment, then the 1st, 5th, 9th, 13th, 17th, and 21st keyframe may be selected. In this embodiment, a fixed number of keyframes are skipped.

[85] In yet another embodiment, the video keyframes to be displayed for a sub-segment may be selected based upon time values associated with the keyframes in the set of keyframes extracted for sub-segment. For example, if the set of video keyframes extracted for a sub-segment comprises 24 keyframes extracted at a sampling rate of 1 second and if six

video keyframes are to be displayed for each sub-segment, then the first frame may be selected and subsequently a keyframe occurring 4 seconds after the previously selected keyframe may be selected.

[86] In an alternative embodiment of the present invention, server 104 may select keyframes from the set of keyframes based upon differences in the contents of the keyframes. For each sub-segment, server 104 may use special image processing techniques to determine differences in the contents of the keyframes extracted for the sub-segment. If six video keyframes are to be displayed for each sub-segment, server 104 may then select six keyframes from the set of extracted keyframes based upon the results of the image processing techniques. For example, the six most dissimilar keyframes may be selected for display in panel 324-2. It should be apparent that various other techniques known to those skilled in the art may also be used to perform the selection of video keyframes.

[87] The selected keyframes are then displayed in panel 324-2. Various different formats may be used to display the selected keyframes in panel 324-2. For example, as shown in Fig. 3, for each sub-segment, the selected keyframes are laid out left-to-right and top-to-bottom.

[88] In an alternative embodiment of the present invention, the entire multimedia document is divided into sub-segments of a pre-determined time period. Each sub-segment is characterized by a start time and an end time associated with the sub-segment. According to an embodiment of the present invention, the start time of the first sub-segment corresponds to the start time of the multimedia document while the end time of the last sub-segment corresponds to the end time of the multimedia document. As described above, server 104 then extracts a set of one or more video keyframes from the video information stored by the multimedia document for each sub-segment based upon the start time and end time associated with the sub-segment. Server 104 then selects one or more keyframes for display for each sub-segment. Based upon the position of thumbnail viewing area lens 314, keyframes that have been selected for display and that occur between  $t_1$  and  $t_2$  associated with thumbnail viewing area lens 314 are then displayed in panel 324-2.

[89] It should be apparent that various other techniques may also be used for displaying video information in panel 324-2 in alternative embodiments of the present invention. According to an embodiment of the present invention, the user may configure the technique to be used for displaying video information in third viewing area 306.

[90] In GUI 300 depicted in Fig. 3, each sub-segment is 8 seconds long and video keyframes corresponding to a plurality of sub-segments are displayed in panel 324-2.

Six video keyframes are displayed from each sub-segment. For each sub-segment, the displayed keyframes are laid out in a left-to-right and top-to-bottom manner.

[91] It should be apparent that, in alternative embodiments of the present invention, the number of panels displayed in third viewing area 306 may be more or less than the number of thumbnail images displayed in second viewing area 304. According to an embodiment of the present invention, the number of panels displayed in third viewing area 306 is user-configurable.

[92] According to the teachings of the present invention, a viewing lens or window 322 (hereinafter referred to as "panel viewing area lens 322") is displayed covering or emphasizing a portion of overview region 306. According to the teachings of the present invention, multimedia information corresponding to the area of third viewing area 306 emphasized by panel viewing area lens 322 is displayed in fourth viewing area 308. A user may change the position of panel viewing area lens 322 by sliding or moving lens 322 along third viewing area 306. In response to a change in the position of panel viewing area lens 322 from a first location in third viewing area 306 to a second location, the multimedia information displayed in fourth viewing area 308 is automatically updated such that the multimedia information displayed in fourth viewing area 308 continues to correspond to the area of third viewing area 306 emphasized by panel viewing area lens 322. Accordingly, a user may use panel viewing area lens 322 to change the multimedia information displayed in fourth viewing area 308.

[93] As described above, a change in the location of panel viewing area lens 322 also causes a change in the location of sub-lens 316 such that the area of second viewing area 304 emphasized by sub-lens 316 continues to correspond to the area of third viewing area 306 emphasized by panel viewing area lens 322. Likewise, as described above, a change in the location of sub-lens 316 also causes a change in the location of panel viewing area lens 322 over third viewing area 306 such that the area of third viewing area 306 emphasized by panel viewing area lens 322 continues to correspond to the changed location of sub-lens 316.

[94] Fig. 5A is a zoomed-in simplified diagram of panel viewing area lens 322 according to an embodiment of the present invention. As depicted in Fig. 5A, panel viewing area lens 322 is bounded by a first edge 326 and a second edge 328. Panel viewing area lens 322 emphasizes an area of third viewing area 306 between edge 326 and edge 328. Based upon the position of panel viewing area lens 322 over third viewing area 306, edge 326 corresponds to specific time "t<sub>3</sub>" in the multimedia document and edge 328 corresponds to a

specific time "t<sub>4</sub>" in the multimedia document where  $t_4 > t_3$  and  $(t_1 \leq t_3 < t_4 \leq t_2)$ . For example, when panel viewing area lens 322 is positioned at the start of third viewing area 306, t<sub>3</sub> may be equal to t<sub>1</sub>, and when panel viewing area lens 322 is positioned at the end of third viewing area 306, t<sub>4</sub> may be equal to t<sub>2</sub>. Accordingly, panel viewing area lens 322 emphasizes a portion of third viewing area 306 between times t<sub>3</sub> and t<sub>4</sub>. According to an embodiment of the present invention, multimedia information corresponding to the time segment between t<sub>3</sub> and t<sub>4</sub> (which is emphasized or covered by panel viewing area lens 322) is displayed in fourth viewing area 308. When the position of panel viewing area lens 322 is changed along third viewing area 306 in response to user input, the information displayed in fourth viewing area 308 may be updated such that the multimedia information displayed in fourth viewing area 308 continues to correspond to the area of third viewing area 306 emphasized by panel viewing area lens 322. Third viewing area 306 thus provides a context and indicates the location of the multimedia information displayed in fourth viewing area 308 within the multimedia document.

[95] According to an embodiment of the present invention, a particular line of text (or one or more words from the last line of text) emphasized by panel viewing area lens 322 may be displayed on a section of lens 322. For example, as depicted in Figs. 5A and 3, the last line of text 330 "Environment is a national" that is emphasized by panel viewing area lens 322 in panel 324-1 is displayed in bolded style on panel viewing area lens 322.

[96] According to an embodiment of the present invention, special features may be attached to panel viewing area lens 322 to facilitate browsing and navigation of the multimedia document. As shown in Fig. 5A, a "play/pause button" 332 and a "lock/unlock button" 334 are provided on panel viewing area lens 322 according to an embodiment of the present invention. Play/Pause button 332 allows the user to control playback of the video information from panel viewing area lens 322. Lock/Unlock button 334 allows the user to switch the location of the video playback from area 340-1 of fourth viewing area 308 to a reduced window on top of panel viewing area lens 322.

[97] Fig. 5B is a simplified example of panel viewing area lens 322 with its lock/unlock button 334 activated or "locked" (i.e., the video playback is locked onto panel viewing area lens 322) according to an embodiment of the present invention. As depicted in Fig. 5B, in the locked mode, the video information is played back on a window 336 on lens 322. In the embodiment depicted in Fig. 5B, the portion of panel viewing area lens 322 over panel 342-2 is expanded in size beyond times t<sub>3</sub> and t<sub>4</sub> to accommodate window 336.

According to an embodiment of the present invention, the video contents displayed in window 336 correspond to the contents displayed in area 340-1 of fourth viewing area 308.

[98] According to an embodiment of the present invention, window 336 has transparent borders so that portions of the underlying third viewing area 306 (e.g., the keyframes displayed in panel 324-2) can be seen. This helps to maintain the user's location focus while viewing third viewing area 306. The user may use play/pause button 332 to start and stop the video displayed in window 336. The user may change the location of panel viewing area lens 322 while the video is being played back in window 336. A change in the location of panel viewing area lens 322 causes the video played back in window 336 to change corresponding to the new location of panel viewing area lens 322. The video played back in window 336 corresponds to the new time values  $t_3$  and  $t_4$  associated with panel viewing area lens 322.

[99] Fig. 5C is a simplified example of panel viewing area lens 322 wherein a representative video keyframe is displayed on panel viewing area lens 322 according to an embodiment of the present invention. In this embodiment server 104 analyzes the video keyframes of panel 324-2 emphasized or covered by panel viewing area lens 322 and determines a particular keyframe 338 that is most representative of the keyframes emphasized by panel viewing area lens 322. The particular keyframe is then displayed on a section of panel viewing area lens 322 covering panel 324-2. In the embodiment depicted in Fig. 5C, the portion of panel viewing area lens 322 over panel 342-2 is expanded in size beyond times  $t_3$  and  $t_4$  to accommodate display of keyframe 338.

[100] As described above, multimedia information corresponding to the section of third viewing area 306 covered by panel viewing area lens 322 (i.e., multimedia information occurring in the time segment between  $t_3$  and  $t_4$ ) is displayed in fourth viewing area 308. As depicted in Fig. 3, fourth viewing area 308 may comprise one or more sub viewing areas 340 (e.g., 340-1, 340-2, and 340-3). According to an embodiment of the present invention, one or more of sub-regions 340 may display a particular type of information included in the multimedia information corresponding to the section of third viewing area 306 emphasized by panel viewing area lens 322.

[101] For example, as depicted in Fig. 3, video information corresponding to (or starting from) the video information emphasized by panel viewing area lens 322 in third viewing area 306 is displayed in sub viewing area 340-1. According to an embodiment of the present invention, video information starting at time  $t_3$  (time corresponding to the top edge of panel viewing area lens 322) may be played back in sub viewing area 340-1. In alternative



embodiments, the video information played back in area 340-1 may start at time  $t_4$  or some other user-configurable time between  $t_3$  and  $t_4$ . The playback of the video in sub viewing area 340-1 may be controlled using control bar 342. Control bar 342 provides a plurality of controls including controls for playing, pausing, stopping, rewinding, and forwarding the video played in sub viewing area 340-1. The current time and length 344 of the video being played in area 340-1 is also displayed. Information identifying the name of the video 346, the date 348 the video was recorded, and the type of the video 350 is also displayed.

[102] In alternative embodiments of the present invention, instead of playing back video information, a video keyframe from the video keyframes emphasized by panel viewing area lens 322 in panel 324-2 is displayed in sub viewing area 340-1. According to an embodiment of the present invention, the keyframe displayed in area 340-1 represents a keyframe that is most representative of the keyframes emphasized by panel viewing area lens 322.

[103] According to an embodiment of the present invention, text information (e.g., CC text, transcript of audio information, etc.) emphasized by panel viewing area lens 322 in third viewing area 306 is displayed in sub viewing area 340-2. According to an embodiment of the present invention, sub viewing area 340-2 displays text information that is displayed in panel 324-1 and emphasized by panel viewing area lens 322. As described below, various types of information may be displayed in sub viewing area 340-3.

[104] Additional information related to the multimedia information stored by the multimedia document may be displayed in fifth viewing area 310 of GUI 300. For example, as depicted in Fig. 3, words occurring in the text information included in the multimedia information displayed by GUI 300 are displayed in area 352 of fifth viewing area 310. The frequency of each word in the multimedia document is also displayed next to each word. For example, the word "question" occurs seven times in the multimedia information CC text. Various other types of information related to the multimedia information may also be displayed in fifth viewing area 310.

[105] According to an embodiment of the present invention, GUI 300 provides features that enable a user to search for one or more words that occur in the text information (e.g., CC text, transcript of audio information) extracted from the multimedia information. For example, a user can enter one or more query words in input field 354 and upon selecting "Find" button 356, server 104 analyzes the text information extracted from the multimedia information stored by the multimedia document to identify all occurrences of the one or more query words entered in field 354. The occurrences of the one or more words in

the multimedia document are then highlighted when displayed in second viewing area 304, third viewing area 306, and fourth viewing area 308. For example, according to an embodiment of the present invention, all occurrences of the query words are highlighted in thumbnail image 312-1, in panel 324-1, and in sub viewing area 340-2. In alternative  
5 embodiments of the present invention, occurrences of the one or more query words may also be highlighted in the other thumbnail images displayed in second viewing area 304, panels displayed in third viewing area 306, and sub viewing areas displayed in fourth viewing area 308.

[106] The user may also specify one or more words to be highlighted in the  
10 multimedia information displayed in GUI 300. For example, a user may select one or more words to be highlighted from area 352. All occurrences of the keywords selected by the user in area 352 are then highlighted in second viewing area 304, third viewing area 306, and fourth viewing area 308. For example, as depicted in Fig. 6, the user has selected the word "National" in area 352. In response to the user's selection, according to an embodiment of the present invention, all occurrences of the word "National" are highlighted in second  
15 viewing area 304, third viewing area 306, and third viewing area 306.

[107] According to an embodiment of the present invention, lines of text 360 that comprise the user-selected word(s) (or query words entered in field 354) are displayed in sub viewing area 340-3 of fourth viewing area 308. For each line of text, the time 362 when the line occurs (or the timestamp associated with the line of text) in the multimedia document is also displayed. The timestamp associated with the line of text generally corresponds to the timestamp associated with the first word in the line.

[108] For each line of text, one or more words surrounding the selected or query word(s) are displayed. According to an embodiment of the present invention, the  
25 number of words surrounding a selected word that are displayed in area 340-3 are user configurable. For example, in GUI 300 depicted in Fig. 6, a user can specify the number of surrounding words to be displayed in area 340-3 using control 364. The number specified by the user indicates the number of words that occur before the select word and the number of words that occur after the selected word that are to be displayed. In the embodiment depicted  
30 in Fig. 6, control 364 is a slider bar that can be adjusted between a minimum value of "3" and a maximum value of "10". The user can specify the number of surrounding words to be displayed by adjusting slider bar 364. For example, if the slider bar is set to "3", then three words that occur before a selected word and three words that occur after the selected word will be displayed in area 340-3. The minimum and maximum values are user configurable.

10081135 022102 2022061801

[109] Further, GUI 300 depicted in Fig. 6 comprises an area 358 sandwiched between thumbnail images 312-1 and 312-2 that indicates locations of occurrences of the query words or other words specified by the user. For example, area 358 comprises markers indicating the locations of word "National" in thumbnail image 312-1. The user can then use either thumbnail viewing area lens 314, or panel viewing area lens 322 to scroll to a desired location within the multimedia document. Fig. 7 depicts a simplified zoomed-in view of second viewing area 304 showing area 358 according to an embodiment of the present invention. As depicted in Fig. 7, area 358 (or channel 358) comprises markers 360 indicating locations in thumbnail image 312-1 that comprise occurrences of the word "National". In alternative embodiments of the present invention, markers in channel 358 may also identify locations of the user-specified words or phrases in the other thumbnail images displayed in second viewing area 304. In alternative embodiments, locations of occurrences of the query words or other words specified by the user may be displayed on thumbnail images 312 (as depicted in Fig. 20A).

[110] As shown in Fig. 6, the position of thumbnail viewing area lens 314 has been changed with respect to Fig. 3. In response to the change in position of thumbnail viewing area lens 314, the multimedia information displayed in third viewing area 306 has been changed to correspond to the section of second viewing area 304 emphasized by thumbnail viewing area lens 314. The multimedia information displayed in fourth viewing area 308 has also been changed corresponding to the new location of panel viewing area lens 322.

[111] According to an embodiment of the present invention, multimedia information displayed in GUI 300 that is relevant to user-specified topics of interest is highlighted or annotated. The annotations provide visual indications of information that is relevant to or of interest to the user. GUI 300 thus provides a convenient tool that allows a user to readily locate portions of the multimedia document that are relevant to the user.

[112] According to an embodiment of the present invention, information specifying topics that are of interest or are relevant to the user may be stored in a user profile. One or more words or phrases may be associated with each topic of interest. Presence of the one or more words and phrases associated with a particular user-specified topic of interest indicates presence of information related to the particular topic. For example, a user may specify two topics of interest--"George W. Bush" and "Energy Crisis". Words or phrases associated with the topic "George Bush" may include "President Bush," "the President," "Mr. Bush," and other like words and phrases. Words or phrases associated with the topic "Energy

Crisis" may include "industrial pollution," "natural pollution," "clean up the sources," "amount of pollution," "air pollution", "electricity," "power-generating plant," or the like. Probability values may be associated with each of the words or phrases indicating the likelihood of the topic of interest given the presence of the word or phrase. Various tools  
5 may be provided to allow the user to configure topics of interest, to specify keywords and phrases associated with the topics, and to specify probability values associated with the keywords or phrases.

[113] It should be apparent that various other techniques known to those skilled in the art may also be used to model topics of interest to the user. These techniques  
10 may include the use of Bayesian networks, relevance graphs, or the like. Techniques for determining sections relevant to user-specified topics, techniques for defining topics of interest, techniques for associating keywords and/or key phrases and probability values are described in U.S. Application No. 08/995,616, filed December 22, 1997, the entire contents of which are herein incorporated by reference for all purposes.

[114] According to an embodiment of the present invention, in order to identify locations in the multimedia document related to user-specified topics of interest, server 104 searches the multimedia document to identify locations within the multimedia document of words or phrases associated with the topics of interest. As described above, presence of words and phrases associated with a particular user-specified topic of interest in the multimedia document indicate presence of the particular topic relevant to the user. The words and phrases that occur in the multimedia document and that are associated with user specified topics of interest are annotated when displayed by GUI 300.  
15  
20

[115] Fig. 8 depicts an example of a simplified GUI 800 in which multimedia information that is relevant to one or more topics of interest to a user is annotated  
25 (or highlighted) when displayed in GUI 800 according to an embodiment of the present invention. GUI 800 depicted in Fig. 8 is merely illustrative of an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[116] In the embodiment depicted in Fig. 8, the user has specified four topics  
30 of interest 802. A label 803 identifies each topic. The topics specified in GUI 800 include "Energy Crisis," "Assistive Tech," "George W. Bush." and "Nepal." In accordance with the teachings of the present invention, keywords and key phrases relevant to the specified topics are highlighted in second viewing area 304, third viewing area 306, and fourth viewing area 308. Various different techniques may be used to highlight or annotate the keywords and/or

key phrases related to the topics of interest. According to an embodiment of the present invention, different colors and styles (e.g., bolding, underlining, different font size, etc.) may be used to highlight words and phrases related to user-specified topics. For example, each topic may be assigned a particular color and content related to a particular topic might be highlighted using the particular color assigned to the particular topic. For example, as depicted in Fig. 8, a first color is used to highlight words and phrases related to the "Energy Crisis" topic of interest, a second color is used to highlight words and phrases related to the "Assistive Tech" topic of interest, a third color is used to highlight words and phrases related to the "George W. Bush" topic of interest, and a fourth color is used to highlight words and phrases related to the "Nepal" topic of interest.

[117] According to an embodiment of the present invention, server 104 searches the text information (either CC text or transcript of audio information) extracted from the multimedia information to locate words or phrases relevant to the user topics. If server 104 finds a word or phrase in the text information that is associated with a topic of interest, the word or phrase is annotated when displayed in GUI 800. As described above, several different techniques may be used to annotate the word or phrase. For example, the word or phrase may be highlighted, bolded, underlined, demarcated using sidebars or balloons, font may be changed, etc.

[118] Keyframes (representing video information of the multimedia document) that are displayed by the GUI and that are related to user specified topics of interest may also be highlighted. According to an embodiment of the present invention, server system 104 may use OCR techniques to extract text from the keyframes extracted from the video information included in the multimedia information. The text output of the OCR techniques may then be compared with words or phrases associated with one or more user-specified topics of interest. If there is a match, the keyframe containing the matched word or phrase (i.e., the keyframe from which the matching word or phrase was extracted by OCR techniques) may be annotated when the keyframe is displayed in GUI 800 either in second viewing area 304, third viewing area 306, or fourth viewing area 308 of GUI 800. Several different techniques may be used to annotate the keyframe. For example, a special box may be drawn around a keyframe that is relevant to a particular topic of interest. The color of the box may correspond to the color associated with the particular topic of interest. The matching text in the keyframe may also be highlighted or underlined or displayed in reverse video. As described above, the annotated keyframes displayed in second viewing area 304 (e.g., the keyframes displayed in thumbnail image 312-2 in Fig. 3) may be identified by

markers displayed in channel area 358. In alternative embodiments, the keyframes may be annotated in thumbnail image 312-2.

[119] According to an embodiment of the present invention, as shown in Fig. 8, a relevance indicator 804 may also be displayed for each user topic. For a particular topic, the relevance indicator for the topic indicates the degree of relevance (or a relevancy score) of the multimedia document to the particular topic. For example, as shown in Fig. 8, the number of bars displayed in a relevance indicator associated with a particular topic indicates the degree of relevance of the multimedia document to the particular topic. Accordingly, the multimedia document displayed in GUI 800 is most relevant to user topic "Energy Crisis" (as indicated by four bars) and least relevant to user topic "Nepal" (indicated by one bar). Various other techniques (e.g., relevance scores, bar graphs, different colors, etc.) may also be used to indicate the degree of relevance of each topic to the multimedia document.

[120] According to an embodiment of the present invention, the relevancy score for a particular topic may be calculated based upon the frequency of occurrences of the words and phrases associated with the particular topic in the multimedia information. Probability values associated with the words or phrases associated with the particular topic may also be used to calculate the relevancy score for the particular topic. Various techniques known to those skilled in the art may also be used to determine relevancy scores for user specified topics of interest based upon the frequency of occurrences of words and phrases associated with a topic in the multimedia information and the probability values associated with the words or phrases. Various other techniques known to those skilled in the art may also be used to calculate the degree of relevancy of the multimedia document to the topics of interest.

[121] As previously stated, a relevance indicator is used to display the degree or relevancy or relevancy score to the user. Based upon information displayed by the relevance indicator, a user can easily determine relevance of multimedia information stored by a multimedia document to topics that may be specified by the user.

[122] Fig. 9 depicts a simplified user interface 900 for defining a topic of interest according to an embodiment of the present invention. User interface 900 may be invoked by selecting an appropriate command from first viewing area 302. GUI 900 depicted in Fig. 9 is merely illustrative of an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

1081129-022102

[123] A user may specify a topic of interest in field 902. A label identifying the topic of interest can be specified in field 910. The label specified in field 910 is displayed in the GUI generated according to the teachings of the present invention to identify the topic of interest. A list of keywords and/or phrases associated with the topic specified in field 902 is displayed in area 908. A user may add new keywords to the list, modify one or more keywords in the list, or remove one or more keywords from the list of keywords associated with the topic of interest. The user may specify new keywords or phrases to be associated with the topic of interest in field 904. Selection of "Add" button 906 adds the keywords or phrases specified in field 904 to the list of keywords previously associated with a topic. The user may specify a color to be used for annotating information relevant to the topic of interest by selecting the color in area 912. For example, in the embodiment depicted in Fig. 9, locations in the multimedia document related to "Assistive Technology" will be annotated in blue color.

[124] According to the teachings of the present invention, various different types of information included in multimedia information may be displayed by the GUI generated by server 104. Fig. 10 depicts a simplified user interface 1000 that displays multimedia information stored by a meeting recording according to an embodiment of the present invention. It should be apparent that GUI 1000 depicted in Fig. 10 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[125] The multimedia information stored by the meeting recording may comprise video information, audio information and possibly CC text information, and slides information. The slides information may comprise information related to slides (e.g., a PowerPoint presentation slides) presented during the meeting. For example, slides information may comprise images of slides presented at the meeting. As shown in Fig. 10, second viewing area 304 comprises three thumbnail images 312-1, 312-2, and 312-3. Text information (either CC text information or a transcript of audio information included in the meeting recording) extracted from the meeting recording multimedia information is displayed in thumbnail image 312-1. Video keyframes extracted from the video information included in the meeting recording multimedia information are displayed in thumbnail image 312-2. Slides extracted from the slides information included in the multimedia information are displayed in thumbnail image 312-3. The thumbnail images are temporally aligned with one another. The information displayed in thumbnail image 312-4 provides additional context for

the video and text information in that, the user can view presentation slides that were presented at various times throughout the meeting recording.

[126] Third viewing area 306 comprises three panels 324-1, 324-2, and 324-

3. Panel 324-1 displays text information corresponding to the section of thumbnail image 312-1 emphasized or covered by thumbnail viewing area lens 314. Panel 324-2 displays video keyframes corresponding to the section of thumbnail image 312-2 emphasized or covered by thumbnail viewing area lens 314. Panel 324-3 displays one or more slides corresponding to the section of thumbnail image 312-3 emphasized or covered by thumbnail viewing area lens 314. The panels are temporally aligned with one another.

[127] Fourth viewing area 308 comprises three sub-viewing areas 340-1, 340-2, and 340-3. Sub viewing area 340-1 displays video information corresponding to the section of panel 324-2 covered by panel viewing area lens 322. As described above, sub-viewing area 340-1 may display a keyframe corresponding to the emphasized portion of panel 324-2. Alternatively, video based upon the position of panel viewing area lens 322 may be played back in area 340-1. According to an embodiment of the present invention, time  $t_3$  associated with lens 322 is used as the start time for playing the video in area 340-1 of fourth viewing area 308. A panoramic shot 1002 of the meeting room (which may be recorded using a 360 degrees camera) is also displayed in area 340-1 of fourth viewing area 308. Text information emphasized by panel viewing area lens 322 in panel 324-1 is displayed in area 340-2 of fourth viewing area 308. One or more slides emphasized by panel viewing area lens 322 in panel 324-3 are displayed in area 340-3 of fourth viewing area 308. According to an embodiment of the present invention, the user may also select a particular slide from panel 324-3 by clicking on the slide. The selected slide is then displayed in area 340-3 of fourth viewing area 308.

[128] According to an embodiment of the present invention, the user can specify the types of information included in the multimedia document that are to be displayed in the GUI. For example, the user can turn on or off slides related information (i.e., information displayed in thumbnail 312-3, panel 324-3, and area 340-3 of fourth viewing area 308) displayed in GUI 1000 by selecting or deselecting "Slides" button 1004. If a user deselects slides information, then thumbnail 312-3 and panel 324-3 are not displayed by GUI 1000. Thumbnail 312-3 and panel 324-3 are displayed by GUI 1000 if the user selects button 1004. Button 1004 thus acts as a switch for displaying or not displaying slides information. In a similar manner, the user can also control other types of information displayed by a GUI generated according to the teachings of the present invention. For example, features may be



provided for turning on or off video information, text information, and other types of information that may be displayed by GUI 1000.

[129] Fig. 11 depicts a simplified user interface 1100 that displays multimedia information stored by a multimedia document according to an embodiment of the present invention. It should be apparent that GUI 1100 depicted in Fig. 11 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[130] The multimedia document whose contents are displayed in GUI 1100 comprises video information, audio information or CC text information, slides information, and whiteboard information. The whiteboard information may comprise images of text and drawings drawn on a whiteboard. As shown in Fig. 11, second viewing area 304 comprises four thumbnail images 312-1, 312-2, 312-3, and 312-4. Text information (either CC text information or a transcript of audio information included in the meeting recording) extracted from the multimedia document is displayed in thumbnail image 312-1. Video keyframes extracted from the video information included in the multimedia document are displayed in thumbnail image 312-2. Slides extracted from the slides information included in the multimedia information are displayed in thumbnail image 312-3. Whiteboard images extracted from the whiteboard information included in the multimedia document are displayed in thumbnail image 312-4. The thumbnail images are temporally aligned with one another.

[131] Third viewing area 306 comprises four panels 324-1, 324-2, 324-3, and 324-4. Panel 324-1 displays text information corresponding to the section of thumbnail image 312-1 emphasized or covered by thumbnail viewing area lens 314. Panel 324-2 displays video keyframes corresponding to the section of thumbnail image 312-2 emphasized or covered by thumbnail viewing area lens 314. Panel 324-3 displays one or more slides corresponding to the section of thumbnail image 312-3 emphasized or covered by thumbnail viewing area lens 314. Panel 324-4 displays one or more whiteboard images corresponding to the section of thumbnail image 312-4 emphasized or covered by thumbnail viewing area lens 314. The panels are temporally aligned with one another.

[132] Fourth viewing area 308 comprises three sub-viewing areas 340-1, 340-2, and 340-3. Area 340-1 displays video information corresponding to the section of panel 324-2 covered by panel viewing area lens 322. As described above, sub-viewing area 340-1 may display a keyframe or play back video corresponding to the emphasized portion of

panel 324-2. According to an embodiment of the present invention, time  $t_3$  (as described above) associated with lens 322 is used as the start time for playing the video in area 340-1 of fourth viewing area 308. A panoramic shot 1102 of the location where the multimedia document was recorded (which may be recorded using a 360 degrees camera) is also displayed in area 340-1 of fourth viewing area 308. Text information emphasized by panel viewing area lens 322 in panel 324-1 is displayed in area 340-2 of fourth viewing area 308. Slides emphasized by panel viewing area lens 322 in panel 324-3 or whiteboard images emphasized by panel viewing area lens 322 in panel 324-4 may be displayed in area 340-3 of fourth viewing area 308. In the embodiment depicted in Fig. 11, a whiteboard image corresponding to the section of panel 324-4 covered by panel viewing area lens 322 is displayed in area 340-3. According to an embodiment of the present invention, the user may also select a particular slide from panel 324-3 or select a particular whiteboard image from panel 324-4 by clicking on the slide or whiteboard image. The selected slide or whiteboard image is then displayed in area 340-3 of fourth viewing area 308.

[133] As described above, according to an embodiment of the present invention, the user can specify the types of information from the multimedia document that are to be displayed in the GUI. For example, the user can turn on or off a particular type of information displayed by the GUI. "WB" button 1104 allows the user to turn on or off whiteboard related information (i.e., information displayed in thumbnail image 312-4, panel 324-4, and area 340-3 of fourth viewing area 308) displayed in GUI 1000.

[134] Fig. 12 depicts a simplified user interface 1200 that displays contents of a multimedia document according to an embodiment of the present invention. It should be apparent that GUI 1200 depicted in Fig. 12 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[135] As depicted in Fig. 12, preview areas 1202 and 1204 are provided at the top and bottom of third viewing area 306. In this embodiment, panel viewing area lens 322 can be moved along third viewing area 306 between edge 1206 of preview area 1202 and edge 1208 of preview area 1204. Preview areas 1202 and 1204 allow the user to preview the contents displayed in third viewing area 306 when the user scrolls the multimedia document using panel viewing area lens 322. For example, as the user is scrolling down the multimedia document using panel viewing area lens 322, the user can see upcoming contents in preview area 1204 and see the contents leaving third viewing area 306 in preview area 1202. If the

user is scrolling up the multimedia document using panel viewing area lens 322, the user can see upcoming contents in preview area 1202 and see the contents leaving third viewing area 306 in preview area 1204. According to an embodiment of the present invention, the size (or length) of each preview region can be changed and customized by the user. For example, in GUI 1200 depicted in Fig. 12, a handle 1210 is provided that can be used by the user to change the size of preview region 1204. According to an embodiment of the present invention, preview areas may also be provided in second viewing area 304.

[136] Fig. 13 depicts a simplified user interface 1300 that displays contents of a multimedia document according to an embodiment of the present invention. It should be apparent that GUI 1300 depicted in Fig. 13 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[137] As depicted in Fig. 13, text information is displayed in panel 324-1 of third viewing area 306 in compressed format, i.e., the white spaces between the text lines have been removed. This enhances the readability of the text information. The lines of text displayed in panel 324-1 are then used to determine the video frames to be displayed in panel 324-2. According to an embodiment of the present invention, a timestamp is associated with each line of text displayed in panel 324-1. The timestamp associated with a line of text represents the time when the text occurred in the multimedia document being displayed by GUI 1300. In one embodiment, the timestamp associated with a line of text corresponds to the timestamp associated with the first word in the line of text. The lines of text displayed in panel 324-1 are then grouped into groups, with each group comprising a pre-determined number of lines.

[138] Video keyframes are then extracted from the video information stored by the multimedia document for each group of lines depending on time stamps associated with lines in the group. According to an embodiment of the present invention, server 104 determines a start time and an end time associated with each group of lines. A start time for a group corresponds to a time associated with the first (or earliest) line in the group while an end time for a group corresponds to the time associated with the last line (or latest) line in the group. In order to determine keyframes to be displayed in panel 324-2 corresponding to a particular group of text lines, server 104 extracts a set of one or more video keyframes from the portion of the video information occurring between the start and end time associated with the particular group. One or more keyframes are then selected from the extracted set of video

keyframes to be displayed in panel 324-2 for the particular group. The one or more selected keyframes are then displayed in panel 324-1 proximal to the group of lines displayed in panel 324-1 for which the keyframes have been extracted.

[139] For example, in Fig. 13, the lines displayed in panel 324-1 are divided into groups wherein each group comprises 4 lines of text. For each group, the time stamp associated with the first line in the group corresponds to the start time for the group while the time stamp associated with the fourth line in the group corresponds to the end time for the group of lines. Three video keyframes are displayed in panel 324-2 for each group of four lines of text displayed in panel 324-1 in the embodiment depicted in Fig. 13. According to an embodiment of the present invention, the three video keyframes corresponding to a particular group of lines correspond to the first, middle, and last keyframe from the set of keyframes extracted from the video information between the start and end times of the particular group. As described above, various other techniques may also be used to select the video keyframes that are displayed in panel 324-2. For each group of lines displayed in panel 324-1, the keyframes corresponding to the group of lines are displayed such that the keyframes are temporally aligned with the group of lines. In the embodiment depicted in Fig. 13, the height of keyframes for a group of lines is approximately equal to the vertical height of the group of lines.

[140] The number of text lines to be included in a group is user configurable. Likewise, the number of video keyframes to be extracted for a particular group of lines is also user configurable. Further, the video keyframes to be displayed in panel 324-2 for each group of lines can also be configured by the user of the present invention.

[141] The manner in which the extracted keyframes are displayed in panel 324-2 is also user configurable. Different techniques may be used to show the relationships between a particular group of lines and video keyframes displayed for the particular group of lines. For example, according to an embodiment of the present invention, a particular group of lines displayed in panel 324-1 and the corresponding video keyframes displayed in panel 324-2 may be color-coded or displayed using the same color to show the relationship. Various other techniques known to those skilled in the art may also be used to show the relationships.

#### **GUI generation technique according to an embodiment of the present invention**

[142] The following section describes techniques for generating a GUI (e.g., GUI 300 depicted in Fig. 3) according to an embodiment of the present invention. For

purposes of simplicity, it is assumed that the multimedia information to be displayed in the GUI comprises video information, audio information, and CC text information. The task of generating GUI 300 can be broken down into the following tasks: (a) displaying thumbnail 312-1 displaying text information extracted from the multimedia information in second viewing area 304; (b) displaying thumbnail 312-2 displaying video keyframes extracted from the video information included in the multimedia information; (c) displaying thumbnail viewing area lens 314 emphasizing a portion of second viewing area 304 and displaying information corresponding to the emphasized portion of second viewing area 304 in third viewing area 306, and displaying panel viewing area lens 322 emphasizing a portion of third viewing area 306 and displaying information corresponding to the emphasized portion of third viewing area 306 in fourth viewing area 308; and (d) displaying information in fifth viewing area 310.

[143] Fig. 14 is a simplified high-level flowchart 1400 depicting a method of displaying thumbnail 312-1 in second viewing area 304 according to an embodiment of the present invention. The method depicted in Fig. 14 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is performed by server 104. The method depicted in Fig. 14 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[144] As depicted in Fig. 14, the method is initiated when server 104 accesses multimedia information to be displayed in the GUI (step 1402). As previously stated, the multimedia information may be stored in a multimedia document accessible to server 104. As part of step 1402, server 104 may receive information (e.g., a filename of the multimedia document) identifying the multimedia document and the location (e.g., a directory path) of the multimedia document. A user of the present invention may provide the multimedia document identification information. Server 104 may then access the multimedia document based upon the provided information. Alternatively, server 104 may receive the multimedia information to be displayed in the GUI in the form of a streaming media signal, a cable signal, etc. from a multimedia information source. Server system 104 may then store the multimedia information signals in a multimedia document and then use the stored document to generate the GUI according to the teachings of the present invention.

[145] Server 104 then extracts text information from the multimedia information accessed in step 1402 (step 1404). If the multimedia information accessed in step 1402 comprises CC text information, then the text information corresponds to CC text information that is extracted from the multimedia information. If the multimedia information accessed in step 1402 does not comprise CC text information, then in step 1404, the audio information included in the multimedia information accessed in step 1402 is transcribed to generate a text transcript for the audio information. The text transcript represents the text information extracted in step 1404.

[146] The text information determined in step 1404 comprises a collection of lines with each line comprising one or more words. Each word has a timestamp associated with it indicating the time of occurrence of the word in the multimedia information. The timestamp information for each word is included in the CC text information. Alternatively, if the text represents a transcription of audio information, the timestamp information for each word may be determined during the audio transcription process.

[147] As part of step 1404, each line is assigned a start time and an end time based upon words that are included in the line. The start time for a line corresponds to the timestamp associated with the first word occurring in the line, and the end time for a line corresponds to the timestamp associated with the last word occurring in the line.

[148] The text information determined in step 1404, including the timing information, is then stored in a memory location accessible to server 104 (step 1406). In one embodiment, a data structure (or memory structure) comprising a linked list of line objects is used to store the text information. Each line object comprises a linked list of words contained in the line. Timestamp information associated with the words and the lines is also stored in the data structure. The information stored in the data structure is then used to generate GUI

[149] Server 104 then determines a length or height (in pixels) of a panel (hereinafter referred to as "the text canvas") for drawing the text information (step 1408). In order to determine the length of the text canvas, the duration ("*duration*") of the multimedia information (or the duration of the multimedia document storing the multimedia document) in seconds is determined. A vertical pixels-per-second of time ("*pps*") value is also defined. The "*pps*" determines the distance between lines of text drawn in the text canvas. The value of *pps* thus depends on how close the user wants the lines of text to be to each other when displayed and upon the size of the font to be used for displaying the text. According to an

embodiment of the present invention, a 5 pps value is specified with a 6 point font. The overall height (in pixels) of the text canvas ("*textCanvasHeight*") is determined as follows:

$$\text{textCanvasHeight} = \text{duration} * \text{pps}$$

For example, if the *duration* of the multimedia information is 1 hour (i.e., 3600 seconds) and for a *pps* value of 5, the height of the text canvas (*textCanvasHeight*) is 18000 pixels (3600\*5).

[150] Multipliers are then calculated for converting pixel locations in the text canvas to seconds and for converting seconds to pixels locations in the text canvas (step 1410). A multiplier "*pix\_m*" is calculated for converting a given time value (in seconds) to a particular vertical pixel location in the text canvas. The *pix\_m* multiplier can be used to determine a pixel location in the text canvas corresponding to a particular time value. The value of *pix\_m* is determined as follows:

$$\text{pix\_m} = \text{textCanvasHeight} / \text{duration}$$

For example, if *duration* = 3600 seconds and *textCanvasHeight* = 18000 pixels, then *pix\_m* = 18000/3600 = 5.

[151] A multiplier "*sec\_m*" is calculated for converting a particular pixel location in the text canvas to a corresponding time value. The *sec\_m* multiplier can be used to determine a time value for a particular pixel location in the text canvas. The value of *sec\_m* is determined as follows:

$$\text{sec\_m} = \text{duration} / \text{textCanvasHeight}$$

For example, if *duration* = 3600 seconds and *textCanvasHeight* = 18000 pixels, then *sec\_m* = 3600/18000 = 0.2.

[152] The multipliers calculated in step 1410 may then be used to convert pixels to seconds and seconds to pixels. For example, the pixel location in the text canvas of an event occurring at time *t* = 1256 seconds in the multimedia information is: 1256 \* *pix\_m* = 1256 \* 5 = 6280 pixels from the top of the text canvas. The number of seconds corresponding to a pixel location *p* = 231 in the text canvas is: 231 \* *sec\_m* = 231 \* 0.2 = 46.2 seconds.

[153] Based upon the height of the text canvas determined in step 1408 and the multipliers generated in step 1410, positional coordinates (horizontal (X) and vertical (Y) coordinates) are then calculated for words in the text information extracted in step 1404 (step 1412). As previously stated, information related to words and lines and their associated timestamps may be stored in a data structure accessible to server 104. The positional coordinate values calculated for each word might also be stored in the data structure.

[154] The Y (or vertical) coordinate ( $W_y$ ) for a word is calculated by multiplying the timestamp ( $W_t$ ) (in seconds) associated with the word by multiplier  $pix\_m$  determined in step 1410. Accordingly:

$$W_y \text{ (in pixels)} = W_t * pix\_m$$

For example, if a particular word has  $W_t = 539$  seconds (i.e., the word occurs 539 seconds into the multimedia information), then  $W_y = 539 * 5 = 2695$  vertical pixels from the top of the text canvas.

[155] The X (or horizontal) coordinate ( $W_x$ ) for a word is calculated based upon the word's location in the line and the width of the previous words in the line. For example if a particular line (L) has four words, i.e., L:  $W_1 W_2 W_3 W_4$ , then

$$W_x \text{ of } W_1 = 0$$

$$W_x \text{ of } W_2 = (W_x \text{ of } W_1) + (\text{Width of } W_1) + (\text{Spacing between words})$$

$$W_x \text{ of } W_3 = (W_x \text{ of } W_2) + (\text{Width of } W_2) + (\text{Spacing between words})$$

$$W_x \text{ of } W_4 = (W_x \text{ of } W_3) + (\text{Width of } W_3) + (\text{Spacing between words})$$

[156] The words in the text information are then drawn on the text canvas in a location determined by the X and Y coordinates calculated for the words in step 1412 (step 1414).

[157] Server 104 then determines a height of thumbnail 312-1 that displays text information in second viewing area 304 of GUI 300 (step 1416). The height of thumbnail 312-1 (*ThumbnailHeight*) depends on the height of the GUI window used to displaying the multimedia information and the height of second viewing area 304 within the GUI window. The value of *ThumbnailHeight* is set such that thumbnail 312-1 fits in the GUI in the second viewing area 304.

[158] Thumbnail 312-1 is then generated by scaling the text canvas such that the height of thumbnail 312-1 is equal to *ThumbnailHeight* and the thumbnail fits entirely within the size constraints of second viewing area 304 (step 1418). Thumbnail 312-1, which represents a scaled version of the text canvas, is then displayed in second viewing area 304 of GUI 300 (step 1420).

[159] Multipliers are then calculated for converting pixel locations in thumbnail 312-1 to seconds and for converting seconds to pixel locations in thumbnail 312-1 (step 1422). A multiplier "*tpix\_m*" is calculated for converting a given time value (in seconds) to a particular pixel location in thumbnail 312-1. Multiplier *tpix\_m* can be used to determine a pixel location in the thumbnail corresponding to a particular time value. The value of *tpix\_m* is determined as follows:



$$tpix\_m = ThumbnailHeight/duration$$

For example, if *duration* = 3600 seconds and *ThumbnailHeight* = 900, then *tpix\_m* = 900/3600 = 0.25

[160] A multiplier "*tsec\_m*" is calculated for converting a particular pixel location in thumbnail 312-1 to a corresponding time value. Multiplier *tsec\_m* can be used to determine a time value for a particular pixel location in thumbnail 312-1. The value of *tsec\_m* is determined as follows:

$$tsec\_m = duration/ThumbnailHeight$$

For example, if *duration* = 3600 seconds and *ThumbnailHeight* = 900, then *tsec\_m* = 3600/900 = 4.

[161] Multipliers *tpix\_m* and *tsec\_m* may then be used to convert pixels to seconds and seconds to pixels in thumbnail 312-1. For example, the pixel location in thumbnail 312-1 of a word occurring at time *t* = 1256 seconds in the multimedia information is:  $1256 * tpix\_m = 1256 * 0.25 = 314$  pixels from the top of thumbnail 312-1. The number of seconds represented by a pixel location *p* = 231 in thumbnail 312-1 is:  $231 * tsec\_m = 231 * 4 = 924$  seconds.

[162] Fig. 15 is a simplified high-level flowchart 1500 depicting a method of displaying thumbnail 312-2, which depicts video keyframes extracted from the video information, in second viewing area 304 of GUI 300 according to an embodiment of the present invention. The method depicted in Fig. 15 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is performed by server 104. The method depicted in Fig. 15 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[163] For purposes of simplicity, it is assumed that thumbnail 312-1 displaying text information has already been displayed according to the flowchart depicted in Fig. 14. As depicted in Fig. 15, server 104 extracts a set of keyframes from the video information included in the multimedia information (step 1502). The video keyframes may be extracted from the video information by sampling the video information at a particular sampling rate. According to an embodiment of the present invention, keyframes are extracted from the video information at a sampling rate of 1 frame per second. Accordingly,

if the duration of the multimedia information is 1 hour (3600 seconds), then 3600 video keyframes are extracted from the video information in step 1502. A timestamp is associated with each keyframe extracted in step 1502 indicating the time of occurrence of the keyframe in the multimedia information.

[164] The video keyframes extracted in step 1502 and their associated timestamp information is stored in a data structure (or memory structure) accessible to server 104 (step 1504). The information stored in the data structure is then used for generating thumbnail 312-2.

[165] The video keyframes extracted in step 1504 are then divided into groups (step 1506). A user-configurable time period ("*groupTime*") is used to divide the keyframes into groups. According to an embodiment of the present invention, *groupTime* is set to 8 seconds. In this embodiment, each group comprises video keyframes extracted within an 8 second time period window. For example, if the duration of the multimedia information is 1 hour (3600 seconds) and 3600 video keyframes are extracted from the video information using a sampling rate of 1 frame per second, then if *groupTime* is set to 8 seconds, the 3600 keyframes will be divided into 450 groups, with each group comprising 8 video keyframes.

[166] A start and an end time are calculated for each group of frames (step 1508). For a particular group of frames, the start time for the particular group is the timestamp associated with the first (i.e., the keyframe in the group with the earliest timestamp) video keyframe in the group, and the end time for the particular group is the timestamp associated with the last (i.e., the keyframe in the group with the latest timestamp) video keyframe in the group.

[167] For each group of keyframes, server 104 determines a segment of pixels on a keyframe canvas for drawing one or more keyframes from the group of keyframes (step 1510). Similar to the text canvas, the keyframe canvas is a panel on which keyframes extracted from the video information are drawn. The height of the keyframe canvas ("*keyframeCanvasHeight*") is the same as the height of the text canvas ("*textCanvasHeight*") described above (i.e.,  $keyframeCanvasHeight = textCanvasHeight$ ). As a result, multipliers *pix\_m* and *sec\_m* (described above) may be used to convert a time value to a pixel location in the keyframe canvas and to convert a particular pixel location in the keyframe canvas to a time value.

[168] The segment of pixels on the keyframe canvas for drawing keyframes from a particular group is calculated based upon the start time and end time associated with

the particular group. The starting vertical (Y) pixel coordinate ("*segmentStart*") and the end vertical (Y) coordinate ("*segmentEnd*") of the segment of pixels in the keyframe canvas for a particular group of keyframes is calculated as follows:

$$\text{segmentStart} = (\text{Start time of group}) * \text{pix\_m}$$

$$\text{segmentEnd} = (\text{End time of group}) * \text{pix\_m}$$

Accordingly, the height of each segment ("*segmentHeight*") in pixels of the text canvas is:

$$\text{segmentHeight} = \text{segmentEnd} - \text{segmentStart}$$

[169] The number of keyframes from each group of frames to be drawn in each segment of pixels on the text canvas is then determined (step 1512). The number of keyframes to be drawn on the keyframe canvas for a particular group depends on the height of the segment ("*segmentHeight*") corresponding to the particular group. If the value of *segmentHeight* is small only a small number of keyframes may be drawn in the segment such that the drawn keyframes are comprehensible to the user when displayed in the GUI. The value of *segmentHeight* depends on the value of *pps*. If *pps* is small, then *segmentHeight* will also be small. Accordingly, a larger value of *pps* may be selected if more keyframes are to be drawn per segment.

[170] According to an embodiment of the present invention, if the *segmentHeight* is equal to 40 pixels and each group of keyframes comprises 8 keyframes, then 6 out of the 8 keyframes may be drawn in each segment on the text canvas. The number of keyframes to be drawn in a segment is generally the same for all groups of keyframes. For example, in the embodiment depicted in Fig. 3, six keyframes are drawn in each segment on the text canvas.

[171] After determining the number of keyframes to be drawn in each segment of the text canvas, for each group of keyframes, server 104 identifies one or more keyframes from keyframes in the group of keyframes to be drawn on the keyframe canvas (step 1514). Various different techniques may be used for selecting the video keyframes to be displayed in a segment for a particular group of frames. According to one technique, if each group of video keyframes comprises 8 keyframes and if 6 video keyframes are to be displayed in each segment on the keyframe canvas, then server 104 may select the first two video keyframes, the middle two video keyframes, and the last two video keyframes from each group of video keyframes to be drawn on the keyframe canvas. As described above, various other techniques may also be used to select one or more keyframes to be displayed from the group of keyframes. For example, the keyframes may be selected based upon the

sequential positions of the keyframes in the group of keyframes, based upon time values associated with the keyframes, or based upon other criteria.

[172] According to another technique, server 104 may use special image processing techniques to determine similarity or dissimilarity between keyframes in each group of keyframes. If six video keyframes are to be displayed from each group, server 104 may then select six keyframes from each group of keyframes based upon the results of the image processing techniques. According to an embodiment of the present invention, the six most dissimilar keyframes in each group may be selected to be drawn on the keyframe canvas. It should be apparent that various other techniques known to those skilled in the art may also be used to perform the selection of video keyframes.

[173] Keyframes from the groups of keyframes identified in step 1514 are then drawn on the keyframe canvas in their corresponding segments (step 1516). Various different formats may be used for drawing the selected keyframes in a particular segment. For example, as shown in Fig. 3, for each segment, the selected keyframes may be laid out left-to-right and top-to-bottom in rows of 3 frames. Various other formats known to those skilled in the art may also be used to draw the keyframes on the keyframe canvas. The size of each individual keyframe drawn on the keyframe canvas depends on the height (*segmentHeight*) of the segment in which the keyframe is drawn and the number of keyframes to be drawn in the segment. As previously stated, the height of a segment depends on the value of *pps*. Accordingly, the size of each individual keyframe drawn on the keyframe canvas also depends on the value of *pps*.

[174] Server 104 then determines a height (or length) of thumbnail 312-2 that displays the video keyframes in GUI 300 (step 1518). According to the teachings of the present invention, the height of thumbnail 312-2 is set to be the same as the height of thumbnail 312-1 that displays text information (i.e., the height of thumbnail 312-2 is set to *ThumbnailHeight*).

[175] Thumbnail 312-2 is then generated by scaling the keyframe canvas such that the height of thumbnail 312-2 is equal to *ThumbnailHeight* and thumbnail 312-2 fits entirely within the size constraints of second viewing area 304 (step 1520). Thumbnail 312-2, which represents a scaled version of the keyframe canvas, is then displayed in second viewing area 304 of GUI 300 (step 1522). Thumbnail 312-2 is displayed in GUI 300 next to thumbnail image 312-1 and is temporally aligned or synchronized with thumbnail 312-1 (as shown in Fig. 3). Accordingly, the top of thumbnail 312-2 is aligned with the top of thumbnail 312-1.

[176] Multipliers are calculated for thumbnail 312-2 for converting pixel locations in thumbnail 312-2 to seconds and for converting seconds to pixel locations in thumbnail 312-2 (step 1524). Since thumbnail 312-2 is the same length as thumbnail 312-1 and is aligned with thumbnail 312-1, multipliers "*tpix\_m*" and "*tsec\_m*" calculated for thumbnail 312-1 can also be used for thumbnail 312-2. These multipliers may then be used to convert pixels to seconds and seconds to pixels in thumbnail 312-2.

[177] According to the method displayed in Fig. 15, the size of each individual video keyframe displayed in thumbnail 312-2 depends, in addition to other criteria, on the length of thumbnail 312-2 and on the length of the video information. Assuming that the length of thumbnail 312-2 is fixed, then the height of each individual video keyframe displayed in thumbnail 312-2 is inversely proportional to the length of the video information. Accordingly, as the length of the video information increases, the size of each keyframe displayed in thumbnail 312-2 decreases. As a result, for longer multimedia documents, the size of each keyframe may become so small that the video keyframes displayed in thumbnail 312-2 are no longer recognizable by the user. To avoid this, various techniques may be used to display the video keyframes in thumbnail 312-2 in a manner that makes thumbnail 312-2 more readable and recognizable by the user.

[178] Fig. 16 is a simplified high-level flowchart 1600 depicting another method of displaying thumbnail 312-2 according to an embodiment of the present invention. The method depicted in Fig. 16 maintains the comprehensibility and usability of the information displayed in thumbnail 312-2 by reducing the number of video keyframes drawn in the keyframe canvas and displayed in thumbnail 312-2. The method depicted in Fig. 16 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is performed by server 104. The method depicted in Fig. 16 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[179] As depicted in Fig. 16, steps 1602, 1604, 1606, and 1608 are the same as steps 1502, 1504, 1506, and 1508, depicted in Fig. 15 and explained above. After step 1608, one or more groups whose video keyframes are to be drawn in the keyframe canvas are then selected from the groups determined in step 1606 (step 1609). Various different

techniques may be used to select the groups in step 1609. According to one technique, the groups determined in step 1606 are selected based upon a "SkipCount" value that is user-configurable. For example, if *SkipCount* is set to 4, then every fifth group (i.e., 4 groups are skipped) is selected in step 1609. The value of *SkipCount* may be adjusted based upon the length of the multimedia information. According to an embodiment of the present invention, the value of *SkipCount* is directly proportional to the length of the multimedia information, i.e., *SkipCount* is set to a higher value for longer multimedia documents.

[180] For each group selected in step 1609, server 104 identifies one or more keyframes from the group to be drawn on the keyframe canvas (step 1610). As described above, various techniques may be used to select keyframes to be drawn on the keyframe canvas.

[181] The keyframe canvas is then divided into a number of equal-sized row portions, where the number of row portions is equal to the number of groups selected in step 1609 (step 1612). According to an embodiment of the present invention, the height of each row portion is approximately equal to the height of the keyframe canvas ("keyframeCanvasHeight") divided by the number of groups selected in step 1609.

[182] For each group selected in step 1609, a row portion of the keyframe canvas is then identified for drawing one or more video keyframes from the group (step 1614). According to an embodiment of the present invention, row portions are associated with groups in chronological order. For example, the first row is associated with a group with the earliest start time, the second row is associated with a group with the second earliest start time, and so on.

[183] For each group selected in step 1609, one or more keyframes from the group (identified in step 1610) are then drawn on the keyframe canvas in the row portion determined for the group in step 1614 (step 1616). The sizes of the selected keyframes for each group are scaled to fit the row portion of the keyframe canvas. According to an embodiment of the present invention, the height of each row portion is more than the heights of the selected keyframes, and height of the selected keyframes is increased to fit the row portion. This increases the size of the selected keyframes and makes them more visible when drawn on the keyframe canvas. In this manner, keyframes from the groups selected in step 1609 are drawn on the keyframe canvas.

[184] The keyframe canvas is then scaled to form thumbnail 312-2 that is displayed in second viewing area 304 according to steps 1618, 1620, and 1622. Since the height of the keyframes drawn on the keyframe canvas is increased according to an

embodiment of the present invention, as described above, the keyframes are also more recognizable when displayed in thumbnail 312-2. Multipliers are then calculated according to step 1624. Steps 1618, 1620, 1622, and 1624 are similar to steps 1518, 1520, 1522, and 1524, depicted in Fig. 15 and explained above. As described above, by selecting a subset of the groups, the number of keyframes to be drawn on the keyframe canvas and displayed in thumbnail 312-2 is reduced. This in turn increases the height of each individual video keyframe displayed in thumbnail 312-2 thus making them more recognizable when displayed.

[185] Fig. 17 is a simplified high-level flowchart 1700 depicting a method of displaying thumbnail viewing area lens 314, displaying information emphasized by thumbnail viewing area lens 314 in third viewing area 306, displaying panel viewing area lens 322, displaying information emphasized by panel viewing area lens 322 in fourth viewing area 308, and displaying information in fifth viewing area 310 according to an embodiment of the present invention. The method depicted in Fig. 17 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is performed by server 104. The method depicted in Fig. 17 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[186] As depicted in Fig. 17, server 104 first determines a height (in pixels) of each panel ("*PanelHeight*") to be displayed in third viewing area 306 of GUI 300 (step 1702). The value of *PanelHeight* depends on the height (or length) of third viewing area 306. Since the panels are to be aligned to each other, the height of each panel is set to *PanelHeight*. According to an embodiment of the present invention, *PanelHeight* is set to the same value as *ThumbnailHeight*. However, in alternative embodiments of the present invention, the value of *PanelHeight* may be different from the value of *ThumbnailHeight*.

[187] A section of the text canvas (generated in the flowchart depicted in Fig. 14) equal to *PanelHeight* is then identified (step 1704). The section of the text canvas identified in step 1704 is characterized by vertical pixel coordinate ( $P_{start}$ ) marking the starting pixel location of the section, and a vertical pixel coordinate ( $P_{end}$ ) marking the ending pixel location of the section.

[188] Time values corresponding to the boundaries of the section of the text canvas identified in step 1704 (marked by pixel locations  $P_{start}$  and  $P_{end}$ ) are then determined

(step 1706). The multiplier  $sec\_m$  is used to calculate the corresponding time values. A time  $t_1$  (in seconds) corresponding to pixel location  $P_{start}$  is calculated as follows:

$$t_1 = P_{start} * sec\_m$$

A time  $t_2$  (in seconds) corresponding to pixel location  $P_{end}$  is calculated as follows:

$$t_2 = P_{end} * sec\_m$$

[189] A section of the keyframe canvas corresponding to the selected section of the text canvas is then identified (step 1708). Since the height of the keyframe canvas is the same as the height of the keyframe canvas, the selected section of the keyframe canvas also lies between pixels locations  $P_{start}$  and  $P_{end}$  in the keyframe canvas corresponding to times  $t_1$  and  $t_2$ .

[190] The portion of the text canvas identified in step 1704 is displayed in panel 324-1 in third viewing area 306 (step 1710). The portion of the keyframe canvas identified in step 1708 is displayed in panel 324-2 in third viewing area 306 (step 1712).

[191] A panel viewing area lens 322 is displayed covering a section of third viewing area 306 (step 1714). Panel viewing area lens 322 is displayed such that it emphasizes or covers a section of panel 324-1 panel and 324-2 displayed in third viewing area 306 between times  $t_3$  and  $t_4$  where ( $t_1 \leq t_3 < t_4 \leq t_2$ ). The top edge of panel viewing area lens 322 corresponds to time  $t_3$  and the bottom edge of panel viewing area lens 322 corresponds to time  $t_4$ . The height of panel viewing area lens 322 (expressed in pixels) is equal to: (Vertical pixel location in the text canvas corresponding to  $t_4$ ) - (Vertical pixel location in the text canvas corresponding to  $t_3$ ). The width of panel viewing area lens 322 is approximately equal to the width of third viewing area 306 (as shown in Fig. 3).

[192] A portion of thumbnail 312-1 corresponding to the section of text canvas displayed in panel 324-1 and a portion of thumbnail 312-2 corresponding to the section of keyframe canvas displayed in panel 324-2 are then determined (step 1716). The portion of thumbnail 312-1 corresponding to the section of the text canvas displayed in panel 324-1 is characterized by vertical pixel coordinate ( $TN_{start}$ ) marking the starting pixel location of the thumbnail portion, and a vertical pixel coordinate ( $TN_{end}$ ) marking the ending pixel location of the thumbnail portion. The multiplier  $tpix\_m$  is used to determine pixel locations  $TN_{start}$  and  $TN_{end}$  as follows:

$$TN_{start} = t_1 * tpix\_m$$

$$TN_{end} = t_2 * tpix\_m$$



Since thumbnails 312-1 and 312-2 are of the same length and are temporally aligned to one another, the portion of thumbnail 312-2 corresponding to the sections of keyframe canvas displayed in panel 324-2 also lies between pixel locations  $TN_{start}$  and  $TN_{end}$  on thumbnail 312-2.

[193] Thumbnail viewing area lens 314 is then displayed covering portions of thumbnails 312-1 and 312-2 corresponding to the section of text canvas displayed in panel 324-1 and the section of keyframe canvas displayed in panel 324-2 (step 1718). Thumbnail viewing area lens 314 is displayed covering portions of thumbnails 312-1 and 312-2 between pixels locations  $TN_{start}$  and  $TN_{end}$  of the thumbnails. The height of thumbnail viewing area lens 314 in pixels is equal to  $(TN_{end} - TN_{start})$ . The width of thumbnail viewing area lens 314 is approximately equal to the width of second viewing area 304 (as shown in Fig. 3).

[194] A portion of second viewing area 304 corresponding to the section of third viewing area 306 emphasized by panel viewing area lens 322 is then determined (step 1720). In step 1720, server 104 determines a portion of thumbnail 312-1 and a portion of thumbnail 312-2 corresponding to the time period between  $t_3$  and  $t_4$ . The portion of thumbnail 312-1 corresponding to the time window between  $t_3$  and  $t_4$  is characterized by vertical pixel coordinate ( $TN_{Sub_{start}}$ ) corresponding to time  $t_3$  and marking the starting vertical pixel of the thumbnail portion, and a vertical pixel coordinate ( $TN_{Sub_{end}}$ ) corresponding to time  $t_4$  and marking the ending vertical pixel location of the thumbnail portion. Multiplier  $tpix\_m$  is used to determine pixel locations  $TN_{Sub_{start}}$  and  $TN_{Sub_{end}}$  as follows:

$$TN_{Sub_{start}} = t_3 * tpix\_m$$

$$TN_{Sub_{end}} = t_4 * tpix\_m$$

Since thumbnails 312-1 and 312-2 are of the same length and are temporally aligned to one another, the portion of thumbnail 312-2 corresponding to the time period between  $t_3$  and  $t_4$  also lies between pixel locations  $TN_{Sub_{start}}$  and  $TN_{Sub_{end}}$  on thumbnail 312-2.

[195] Sub-lens 316 is then displayed covering portions of thumbnails 312-1 and 312-2 corresponding to the time window between  $t_3$  and  $t_4$  (i.e., corresponding to the portion of third viewing area 306 emphasized by panel viewing area lens 322) (step 1722). Sub-lens 316 is displayed covering portions of thumbnails 312-1 and 312-2 between pixels locations  $TN_{Sub_{start}}$  and  $TN_{Sub_{end}}$ . The height of sub-lens 316 in pixels is equal to  $(TN_{Sub_{end}} - TN_{Sub_{start}})$ . The width of sub-lens 316 is approximately equal to the width of second viewing area 304 (as shown in Fig. 3).

[196] Multimedia information corresponding to the portion of third viewing area 306 emphasized by panel viewing area lens 322 is displayed in fourth viewing area 308 (step 1724). For example, video information starting at time  $t_3$  is played back in area 340-1 of fourth viewing area 308 in GUI 300. In alternative embodiments, the starting time of the video playback may be set to any time between and including  $t_3$  and  $t_4$ . Text information corresponding to the time window between  $t_3$  and  $t_4$  is displayed in area 340-2 of fourth viewing area 308.

[197] The multimedia information may then be analyzed and the results of the analysis are displayed in fifth viewing area 310 (step 1726). For example, the text information extracted from the multimedia information may be analyzed to identify words that occur in the text information and the frequency of individual words. The words and their frequency may be printed in fifth viewing area 310 (e.g., information printed in area 352 of fifth viewing area 310 as shown in Fig. 3). As previously described, information extracted from the multimedia information may be stored in data structures accessible to server 104. For example, text information and video keyframes information extracted from the multimedia information may be stored in one or more data structures accessible to server 104. Server 104 may use the information stored in these data structures to analyze the multimedia information.

### **Multimedia Information Navigation**

[198] As previously described, a user of the present invention may navigate and scroll through the multimedia information stored by a multimedia document and displayed in GUI 300 using thumbnail viewing area lens 314 and panel viewing area lens 322. For example, the user can change the location of thumbnail viewing area lens 314 by moving thumbnail viewing area lens 314 along the length of second viewing area 304. In response to a change in the position of thumbnail viewing area lens 314 from a first location in second viewing area 304 to a second location along second viewing area 304, the multimedia information displayed in third viewing area 306 is automatically updated such that the multimedia information displayed in third viewing area 306 continues to correspond to the area of second viewing area 304 emphasized by thumbnail viewing area lens 314 in the second location.

[199] Likewise, the user can change the location of panel viewing area lens 322 by moving panel viewing area lens 322 along the length of third viewing area 306. In response to a change in the location of panel viewing area lens 322, the position of sub-lens

316 and also possibly thumbnail viewing area lens 314 are updated to continue to correspond to new location of panel viewing area lens 322. The information displayed in fourth viewing area 308 is also updated to correspond to the new location of panel viewing area lens 322.

[200] Fig. 18 is a simplified high-level flowchart 1800 depicting a method of automatically updating the information displayed in third viewing area 306 in response to a change in the location of thumbnail viewing area lens 314 according to an embodiment of the present invention. The method depicted in Fig. 18 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is performed by server 104. The method depicted in Fig. 18 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[201] As depicted in Fig. 18, the method is initiated when server 104 detects a change in the position of thumbnail viewing area lens 314 from a first position to a second position over second viewing area 304 (step 1802). Server 104 then determines a portion of second viewing area 304 emphasized by thumbnail viewing area lens 314 in the second position (step 1804). As part of step 1804, server 104 determines pixel locations ( $TN_{start}$  and  $TN_{end}$ ) in thumbnail 312-1 corresponding to the edges of thumbnail viewing area lens 314 in the second position.  $TN_{start}$  marks the starting vertical pixel location in thumbnail 312-1, and  $TN_{end}$  marks the ending vertical pixel location in thumbnail 312-1. Since thumbnails 312-1 and 312-2 are of the same length and are temporally aligned to one another, the portion of thumbnail 312-2 corresponding to second position of thumbnail viewing area lens 314 also lies between pixel locations  $TN_{start}$  and  $TN_{end}$ .

[202] Server 104 then determines time values corresponding to the second position of thumbnail viewing area lens 314 (step 1806). A time value  $t_1$  is determined corresponding to pixel location  $TN_{start}$  and a time value  $t_2$  is determined corresponding to pixel location  $TN_{end}$ . The multiplier  $tsec\_m$  is used to determine the time values as follows:

$$t_1 = TN_{start} * tsec\_m$$

$$t_2 = TN_{end} * tsec\_m$$

[203] Server 104 then determines pixel locations in the text canvas and the keyframe canvas corresponding to the time values determined in step 1806 (step 1808). A pixel location  $P_{start}$  in the text canvas is calculated based upon time  $t_1$ , and a pixel location

$P_{end}$  in the text canvas is calculated based upon time  $t_2$ . The multiplier  $pix\_m$  is used to determine the locations as follows:

$$P_{start} = t_1 * tpix\_m$$

$$P_{end} = t_2 * tpix\_m$$

- 5 Since the text canvas and the keyframe canvas are of the same length, time values  $t_1$  and  $t_2$  correspond to pixel locations  $P_{start}$  and  $P_{end}$  in the keyframe canvas.

[204] A section of the text canvas between pixel locations  $P_{start}$  and  $P_{end}$  is displayed in panel 324-1 (step 1810). The section of the text canvas displayed in panel 324-1 corresponds to the portion of thumbnail 312-1 emphasized by thumbnail viewing area lens 314 in the second position.

[205] A section of the keyframe canvas between pixel locations  $P_{start}$  and  $P_{end}$  is displayed in panel 324-2 (step 1812). The section of the keyframe canvas displayed in panel 324-2 corresponds to the portion of thumbnail 312-2 emphasized by thumbnail viewing area lens 314 in the second position.

[206] When thumbnail viewing area lens 314 is moved from the first position to the second position, sub-lens 316 also moves along with thumbnail viewing area lens 314. Server 104 then determines a portion of second viewing area 304 emphasized by sub-lens 316 in the second position (step 1814). As part of step 1814, server 104 determines pixel locations ( $TNSub_{start}$  and  $TNSub_{end}$ ) in thumbnail 312-1 corresponding to the edges of sub-lens 316 in the second position.  $TNSub_{start}$  marks the starting vertical pixel location in thumbnail 312-1, and  $TNSub_{end}$  marks the ending vertical pixel location of sub-lens 316 in thumbnail 312-1. Since thumbnails 312-1 and 312-2 are of the same length and are temporally aligned to one another, the portion of thumbnail 312-2 corresponding to second position of sub-lens 316 also lies between pixel locations  $TNSub_{start}$  and  $TNSub_{end}$ .

[207] Server 104 then determines time values corresponding to the second position of sub-lens 316 (step 1816). A time value  $t_3$  is determined corresponding to pixel location  $TNSub_{start}$  and a time value  $t_4$  is determined corresponding to pixel location  $TNSub_{end}$ . The multiplier  $tsec\_m$  is used to determine the time values as follows:

$$t_3 = TNSub_{start} * tsec\_m$$

$$t_4 = TNSub_{end} * tsec\_m$$

[208] Server 104 then determines pixel locations in the text canvas and the keyframe canvas corresponding to the time values determined in step 1816 (step 1818). A pixel location  $PSub_{start}$  in the text canvas is calculated based upon time  $t_3$ , and a pixel location

$PSub_{end}$  in the text canvas is calculated based upon time  $t_4$ . The multiplier  $pix\_m$  is used to determine the locations as follows:

$$PSub_{start} = t_3 * tpix\_m$$

$$PSub_{end} = t_4 * tpix\_m$$

- 5 Since the text canvas and the keyframe canvas are of the same length, time values  $t_1$  and  $t_2$  correspond to pixel locations  $PSub_{start}$  and  $PSub_{end}$  in the keyframe canvas.

[209] Panel viewing area lens 322 is drawn over third viewing area 306 covering a portion of third viewing area 306 between pixels location  $PSub_{start}$  and  $PSub_{end}$  (step 1820). The multimedia information displayed in fourth viewing area 308 is then updated to correspond to the new position of panel viewing area lens 322 (step 1822).

10 [210] Fig. 19 is a simplified high-level flowchart 1900 depicting a method of automatically updating the information displayed in fourth viewing area 308 and the positions of thumbnail viewing area lens 314 and sub-lens 316 in response to a change in the location of panel viewing area lens 322 according to an embodiment of the present invention. The method depicted in Fig. 19 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is performed by server 104. The method depicted in Fig. 19 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

As depicted in Fig. 19, the method is initiated when server 104 detects a change in the position of panel viewing area lens 322 from a first position to a second position over third viewing area 306 (step 1902). Server 104 then determines time values corresponding to the second position of panel viewing area lens 322 (step 1904). In step 1904, server 104 determines the pixel locations of the top and bottom edges of panel viewing area lens 322 in the second position. Multiplier  $sec\_m$  is then used to covert the pixel locations to time values. A time value  $t_3$  is determined corresponding to top edge of panel viewing area lens 322 in the second position, and a time value  $t_4$  is determined corresponding to bottom edge of panel viewing area lens 322.

$$t_3 = (\text{Pixel location of top edge of panel viewing area lens 322}) * sec\_m$$

$$t_4 = (\text{Pixel location of bottom edge of panel viewing area lens 322}) * sec\_m$$

10081130010  
2012062100010  
[211] Server 104 then determines pixel locations in second viewing area 304 corresponding to the time values determined in step 1904 (step 1906). A pixel location  $TNSub_{start}$  in a thumbnail (either 312-1 or 312-2 since they aligned and of the same length) in second viewing area 304 is calculated based upon time  $t_3$ , and a pixel location  $TNSub_{end}$  in the thumbnail is calculated based upon time  $t_4$ . The multiplier  $tpix\_m$  is used to determine the locations as follows:

$$TNSub_{start} = t_3 * tpix\_m$$

$$TNSub_{end} = t_4 * tpix\_m$$

[212] Sub-lens 316 is then updated to emphasize a portion of thumbnails 312 in second viewing area 304 between pixel locations determined in step 1906 (step 1908). As part of step 1908, the position of thumbnail viewing area lens 314 may also be updated if pixels positions  $TNSub_{start}$  or  $TNSub_{end}$  lie beyond the boundaries of thumbnail viewing area lens 314 when panel viewing area lens 322 was in the first position. For example, if a user uses panel viewing area lens 322 to scroll third viewing area 306 beyond the *PanelHeight*, then the position of thumbnail viewing area lens 314 is updated accordingly. If the second position of panel viewing area lens 322 lies within *PanelHeight*, then only sub-lens 316 is moved to correspond to the second position of panel viewing area lens 322 and thumbnail viewing area lens 314 is not moved.

[213] As described above, panel viewing area lens 322 may be used to scroll the information displayed in third viewing area 306. For example, a user may move panel viewing area lens 322 to the bottom of third viewing area 306 and cause the contents of third viewing area 306 to be automatically scrolled upwards. Likewise, the user may move panel viewing area lens 322 to the top of third viewing area 306 and cause the contents of third viewing area 306 to be automatically scrolled downwards. The positions of thumbnail viewing area lens 314 and sub-lens 316 are updated as scrolling occurs.

[214] Multimedia information corresponding to the second position of panel viewing area lens 322 is then displayed in fourth viewing area 308 (step 1910). For example, video information corresponding to the second position of panel viewing area lens 322 is displayed in area 340-1 of fourth viewing area 308 and text information corresponding to the second position of panel viewing area lens 322 is displayed in area 340-2 of third viewing area 306.

[215] According to an embodiment of the present invention, in step 1910, server 104 selects a time "t" having a value equal to either  $t_3$  or  $t_4$  or some time value between  $t_3$  and  $t_4$ . Time "t" may be referred to as the "location time". The location time may be user-

configurable. According to an embodiment of the present invention, the location time is set to  $t_4$ . The location time is then used as the starting time for playing back video information in area 340-1 of fourth viewing area 308.

[216] According to an embodiment of the present invention, GUI 300 may operate in two modes: a "full update" mode and a "partial update" mode. The user of the GUI may select the operation mode of the GUI.

[217] When GUI 300 is operating in "full update" mode, the positions of thumbnail viewing area lens 314 and panel viewing area lens 322 are automatically updated to reflect the position of the video played back in area 340-1 of fourth viewing area 308. Accordingly, in "full update" mode, thumbnail viewing area lens 314 and panel viewing area lens 322 keep up or reflect the position of the video played in fourth viewing area 308. The video may be played forwards or backwards using the controls depicted in area 342 of fourth viewing area 308, and the positions of thumbnail viewing area lens 314 and panel viewing area lens 322 change accordingly. The multimedia information displayed in panels 324 in third viewing area 306 is also automatically updated (shifted upwards) to correspond to the position of thumbnail viewing area lens 314 and reflect the current position of the video.

[218] When GUI 300 is operating in "partial update" mode, the positions of thumbnail viewing area lens 314 and panel viewing area lens 322 are not updated to reflect the position of the video played back in area 340-1 of fourth viewing area 308. In this mode, the positions of thumbnail viewing area lens 314 and panel viewing area lens 322 remain static as the video is played in area 340-1 of fourth viewing area 308. Since the position of thumbnail viewing area lens 314 does not change, the multimedia information displayed in third viewing area 306 is also not updated. In this mode, a "location pointer" may be displayed in second viewing area 304 and third viewing area 306 to reflect the current position of the video played back in area 340-1 of fourth viewing area 308. The position of the location pointer is continuously updated to reflect the position of the video.

### Ranges

[219] According to an embodiment, the present invention provides techniques for selecting or specifying portions of the multimedia information displayed in the GUI. Each portion is referred to as a "range." A range may be manually specified by a user of the present invention or may alternatively be automatically selected by the present invention based upon range criteria provided by the user of the invention.

[220] A range refers to a portion of the multimedia information between a start time ( $R_S$ ) and an end time ( $R_E$ ). Accordingly, each range is characterized by an  $R_S$  and a  $R_E$  that define the time boundaries of the range. A range comprises a portion of the multimedia information occurring between times  $R_S$  and  $R_E$  associated with the range.

[221] Fig. 20A depicts a simplified user interface 2000 that displays ranges according to an embodiment of the present invention. It should be apparent that GUI 2000 depicted in Fig. 20A is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[222] As depicted in Fig. 20A, GUI 2000 provides various features (buttons, tabs, etc.) that may be used by the user to either manually specify one or more ranges or to configure GUI 2000 to automatically generate ranges. In the embodiment depicted in Fig. 20A, the user can manually specify a range by selecting "New" button 2002. After selecting button 2002, the user can specify a range by selecting a portion of a thumbnail displayed in second viewing area 2004. One or more ranges may be specified by selecting various portions of the thumbnail. For example, in Fig. 20A, six ranges 2006-1, 2006-2, 2006-3, 2006-4, 2006-5, and 2006-6 have been displayed. One or more of these ranges may be manually specified by the user by selecting or marking portions of thumbnail 2008-2. In Fig. 20A, each specified range is indicated by a bar displayed over thumbnail 2008-2. An identifier or label may also be associated with each range to uniquely identify the range. In Fig. 20A, each range is identified by a number associated with the range and displayed in the upper left corner of the range. The numbers act as labels for the ranges.

[223] Each range specified by selecting a portion of thumbnail 2008-2 is bounded by a top edge ( $R_{top}$ ) and a bottom edge ( $R_{bottom}$ ). The  $R_S$  and  $R_E$  times for a range may be determined from the pixel locations of  $R_{top}$  and  $R_{bottom}$  as follows:

$$R_S = R_{top} * tsec\_m$$

$$R_E = R_{bottom} * tsec\_m$$

[224] It should be apparent that various other techniques may also be used for specifying a range. For example, in alternative embodiments of the present invention, a user may specify a range by providing the start time ( $R_S$ ) and end time ( $R_E$ ) for the range.

[225] In GUI 2000 depicted Fig. 20A, information related to the ranges displayed is GUI 2000 is displayed in area 2010. The information displayed for each range in area 2010 includes a label or identifier 2012 identifying the range, a start time ( $R_S$ ) 2014 of the range, an end time ( $R_E$ ) 2016 of the range, a time span 2018 of the range, and a set of



video keyframes 2019 extracted from the portion of the multimedia information associated with the range. The time span for a ranges is calculated by determining the difference between the end time  $R_E$  and the start time associated with the range (i.e., time span for a range =  $R_E - R_S$ ). In the embodiment depicted in Fig. 20A, the first, last, and middle keyframe extracted from the multimedia information corresponding to each range are displayed. Various other techniques may also be used for selecting keyframes to be displayed for a range. The information depicted in Fig. 20A is not meant to limit the scope of the present invention. Various other types of information for a range may also be displayed in alternative embodiments of the present invention.

[226] According to the teachings of the present invention, various operations may be performed on the ranges displayed in GUI 2000. A user can edit a range by changing the  $R_S$  and  $R_E$  times associated with the range. Editing a range may change the time span (i.e., the value of  $(R_E - R_S)$ ) of the range. In GUI 2000 depicted in Fig. 20A, the user can modify or edit a displayed range by selecting "Edit" button 2020. After selecting "Edit" button 2020, the user can edit a particular range by dragging the top edge and/or the bottom edge of the bar representing the range. A change in the position of top edge modifies the start time ( $R_S$ ) of the range, and a change in the position of the bottom edge modifies the end time ( $R_E$ ) of the range.

[227] The user can also edit a range by selecting a range in area 2010 and then selecting "Edit" button 2020. In this scenario, selecting "Edit" button 2020 causes a dialog box to be displayed to the user (e.g., dialog box 2050 depicted in Fig. 20B). The user can then change the  $R_S$  and  $R_E$  values associated with the selected range by entering the values in fields 2052 and 2054, respectively. The time span of the selected range is displayed in area 2056 of the dialog box.

[228] The user can also move the location of a displayed range by changing the position of the displayed range along thumbnail 2008-2. Moving a range changes the  $R_S$  and  $R_E$  values associated with the range but maintains the time span of the range. In GUI 2000, the user can move a range by first selecting "Move" button 2022 and then selecting and moving a range. As described above, the time span for a range may be edited by selecting "Edit" button and then dragging an edge of the bar representing the range.

[229] The user can remove or delete a previously specified range. In GUI 2000 depicted in Fig. 20A, the user can delete a displayed range by selecting "Remove" button 2024 and then selecting the range that is to be deleted. Selection of "Clear" button

2026 deletes all the ranges that have been specified for the multimedia information displayed in GUI 2000.

[230] As indicated above, each range refers to a portion of the multimedia information occurring between times  $R_S$  and  $R_E$  associated with the range. The multimedia information corresponding to a range may be output to the user by selecting "Play" button 2028. After selecting "Play" button 2028, the user may select a particular range displayed in GUI 2000 whose multimedia information is to be output to the user. The portion of the multimedia information corresponding to the selected range is then output to the user. Various different techniques known to those skilled in the art may be used to output the multimedia information to the user. According to an embodiment of the present invention, video information corresponding to multimedia information associated with a selected range is played back to the user in area 2030. Text information corresponding to the selected range may be displayed in area 2032. The positions of thumbnail viewing area lens 314 and panel viewing area lens 322, and the information displayed in third viewing area 306 are automatically updated to correspond to the selected range whose information is output to the user in area 2030.

[231] The user can also select a range in area 2010 and then play information corresponding to the selected range by selecting "Play" button 2020. Multimedia information corresponding to the selected range is then displayed in area 2030.

[232] The user may also instruct GUI 2000 to sequentially output information associated with all the ranges specified for the multimedia information displayed by GUI 2000 by selecting "Preview" button 2034. Upon selecting "Preview" button 2034, multimedia information corresponding to the displayed ranges is output to the user in sequential order. For example, if six ranges have been displayed as depicted in Fig. 20A, multimedia information corresponding to the range identified by label "1" may be output first, followed by multimedia information corresponding to the range identified by label "2", followed by multimedia information corresponding to the range identified by label "3", and so on until multimedia information corresponding to all six ranges has been output to the user. The order in which the ranges are output to the user may be user-configurable.

[233] Multimedia information associated with a range may also be saved to memory. For example, in the embodiment depicted in Fig. 20A, the user may select "Save" button 2036 and then select one or more ranges that are to be saved. Multimedia information corresponding to the ranges selected by the user to be saved is then saved to memory (e.g., a hard disk, a storage unit, a floppy disk, etc.)

[234] Various other operations may also be performed on a range. For example, according to an embodiment of the present invention, multimedia information corresponding to one or more ranges may be printed on a paper medium. Details describing techniques for printing multimedia information on a paper medium are discussed in U.S. Application No. 10/001,895, (Attorney Docket No.: 15358-006500US) filed November 19, 2001, the entire contents of which are herein incorporated by reference for all purposes.

[235] Multimedia information associated with a range may also be communicated to a user-specified recipient. For example, a user may select a particular range and request communication of multimedia information corresponding to the range to a user-specified recipient. The multimedia information corresponding to the range is then communicated to the recipient. Various different communication techniques known to those skilled in the art may be used to communicate the range information to the recipient including faxing, electronic mail, wireless communication, and other communication techniques.

[236] Multimedia information corresponding to a range may also be provided as input to another application program such as a search program, a browser, a graphics application, a MIDI application, or the like. The user may select a particular range and then identify an application to which the information is to be provided. In response to the user's selection, multimedia information corresponding to the range is then provided as input to the application.

[237] As previously stated, ranges may be specified manually by a user or may be selected automatically by the present invention. The automatic selection of ranges may be performed by software modules executing on server 104, hardware modules coupled to server 104, or combinations thereof. Fig. 21 is a simplified high-level flowchart depicting a method of automatically creating ranges according to an embodiment of the present invention. The method depicted in Fig. 21 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is performed by server 104. The method depicted in Fig. 21 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[238] As depicted in Fig. 21, the method is initiated when server 104 receives criteria for creating ranges (step 2102). The user of the present invention may

specify the criteria via GUI 2000. For example, in GUI 2000 depicted in Fig. 20A, area 2040 displays various options that can be selected by the user to specify criteria for automatic creation of ranges. In GUI 2000 depicted in Fig. 20A, the user may select either "Topics" or "Words" as the range criteria. If the user selects "Topics", then information related to topics of interest to the user (displayed in area 2042) is identified as the range creation criteria. If the user selects "Words", then one or more words selected by the user in area 2044 of GUI 2000 are identified as criteria for automatically creating ranges. In alternative embodiments, the criteria for automatically creating ranges may be stored in a memory location accessible to server 104. For example, the criteria information may be stored in a file accessible to server 104. Various other types of criteria may also be specified according to the teachings of the present invention.

[239] The multimedia information stored in the multimedia document is then analyzed to identify locations (referred to as "hits") in the multimedia information that satisfy the criteria received in step 2102 (step 2104). For example, if the user has specified that one or more words selected by the user in area 2044 are to be used as the range creation criteria, then the locations of the selected words are identified in the multimedia information. Likewise, if the user has specified topics of interest as the range creation criteria, then server 104 analyzes the multimedia information to identify locations in the multimedia information that are relevant to the topics of interest specified by the user. As described above, server 104 may analyze the multimedia information to identify locations of words or phrases associated with the topics of interest specified by the user. Information related to the topics of interest may be stored in a user profile file that is accessible to server 104. It should be apparent that various other techniques known to those skilled in the art may also be used to identify locations in the multimedia information that satisfy the range criteria received in step 2102.

[240] One or more ranges are then created based upon the locations of the hits identified in step 2104 (step 2106). Various different techniques may be used to form ranges based upon locations of the hits. According to one technique, one or more ranges are created based upon the times associated with the hits. Hits may be grouped into ranges based on the proximity of the hits to each other. One or more ranges created based upon the locations of the hits may be combined to form larger ranges.

[241] The ranges created in step 2106 are then displayed to the user using GUI 2000 (step 2108). Various different techniques may be used to display the ranges to the user. In Fig. 20A, each range is indicated by a bar displayed over thumbnail 2008-2.

10081121-022106

5 [242] Fig. 22 is a simplified high-level flowchart 2200 depicting a method of automatically creating ranges based upon locations of hits in the multimedia information according to an embodiment of the present invention. The processing depicted in Fig. 22 may be performed in step 2106 depicted in Fig. 21. The method depicted in Fig. 22 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is performed by server 104. The method depicted in Fig. 22 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

5 [243] As depicted in Fig. 22, the method is initiated by determining a time associated with the first hit in the multimedia information (step 2202). The first hit in the multimedia information corresponds to a hit with the earliest time associated with it (i.e., a hit that occurs before other hits in the multimedia information). A new range is then created to include the first hit such that  $R_S$  for the new range is set to the time of occurrence of the first hit, and  $R_E$  for the new range is set to some time value after the time of occurrence of the first hit (step 2204). According to an embodiment of the present invention,  $R_E$  is set to the time of occurrence of the hit plus 5 seconds.

20 [244] Server 104 then determines if there are any additional hits in the multimedia information (step 2206). Processing ends if there are no additional hits in the multimedia information. The ranges created for the multimedia information may then be displayed to the user according to step 2108 depicted in Fig. 21. If it is determined in step 2206 that additional hits exist in the multimedia information, then the time associated with the next hit is determined (step 2208).

25 [245] Server 104 then determines if the time gap between the end time of the range including the previous hit and the time determined in step 2208 exceeds a threshold value (step 2210). Accordingly, in step 2210 server 104 determines if:

30 (Time determined in step 2208) - ( $R_E$  of range including previous hit) > *GapBetweenHits*  
wherein, *GapBetweenHits* represents the threshold time value. The threshold value is user configurable. According to an embodiment of the present invention, *GapBetweenHits* is set to 60 seconds.

[246] If it is determined in step 2210 that the time gap between the end time of the range including the previous hit and the time determined in step 2208 exceeds the

threshold value, then a new range is created to include the next hit such that  $R_S$  for the new range is set to the time determined in step 2208, and  $R_E$  for the new range is set to some time value after the time determined in step 2208 (step 2212). According to an embodiment of the present invention,  $R_E$  is set to the time of occurrence of the hit plus 5 seconds. Processing then continues with step 2206.

[247] If it is determined in step 2210 that the time gap between the end time of the range including the previous hit and the time determined in step 2208 does not exceed the threshold value, then the range including the previous hit is extended by changing the end time  $R_E$  of the range to the time determined in step 2208 (step 2214). Processing then continues with step 2206.

[248] According to the method depicted in Fig. 22, a single range is created for hits in the multimedia information that occur within a threshold value ("GapBetweenHits") from the previous range. At the end of the method depicted in Fig. 22, one or more ranges are automatically created based upon the range criteria.

[249] According to an embodiment of the present invention, after forming one or more ranges based upon the times associated with the hits (e.g., according to flowchart 2200 depicted in Fig. 22), one or more ranges created based upon the locations of the hits may be combined with other ranges to form larger ranges. According to an embodiment of the present invention, a small range is identified and combined with a neighboring range if the time gap between the small range and the neighboring range is within a user-configurable time period threshold. If there are two neighboring time ranges that are within the time period threshold, then the small range is combined with the neighboring range that is closest to the small range. The neighboring ranges do not need to be small ranges. Combination of smaller ranges to form larger ranges is based upon the premise that a larger range is more useful to the user than multiple small ranges.

[250] Fig. 23 is a simplified high-level flowchart 2300 depicting a method of combining one or more ranges based upon the size of the ranges and the proximity of the ranges to neighboring ranges according to an embodiment of the present invention. The processing depicted in Fig. 23 may be performed in step 2106 depicted in Fig. 21 after processing according to flowchart 2200 depicted in Fig. 22 has been performed. The method depicted in Fig. 23 may be performed by server 104, by client 102, or by server 104 and client 102 in combination. For example, the method may be executed by software modules executing on server 104 or on client 102, by hardware modules coupled to server 104 or to client 102, or combinations thereof. In the embodiment described below, the method is

performed by server 104. The method depicted in Fig. 23 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[251] In order to describe the processing performed in Fig. 23, it is assumed that "N" ranges ( $N \geq 1$ ) have been created for the multimedia information displayed by the GUI. The ranges may have been created according to the processing depicted in flowchart 2200 in Fig. 22. Each range  $R_i$ , where ( $1 \leq i \leq N$ ), in the set of "N" ranges has a start time  $R_S$  and an end time  $R_E$  associated with it. For a range  $R_i$ , the neighbors of the range include range  $R_{(i-1)}$  and range  $R_{(i+1)}$ , where  $R_E$  of range  $R_{(i-1)}$  occurs before  $R_S$  of range  $R_i$  and  $R_E$  of range  $R_i$  occurs before  $R_S$  of range  $R_{(i+1)}$ . Range  $R_{(i-1)}$  is referred to as a range that occurs before range  $R_i$ . Range  $R_{(i+1)}$  is referred to as a range that occurs after range  $R_i$ .

[252] As depicted in Fig. 23, the method is initiated by initializing a variable "i" to 1 (step 2303). A range  $R_i$  is then selected (step 2304). During the first pass through flowchart 2300, the first range (i.e., the range having the earliest  $R_S$  time) in the set of "N" ranges is selected. Subsequent ranges are selected in subsequent passes.

[253] Server 104 then determines if range  $R_i$  selected in step 2304 qualifies as a small range. According to an embodiment of the present invention, a threshold value "SmallRangeSize" is defined and a range is considered a small range if the time span of the range is less than or equal to threshold value *SmallRangeSize*. Accordingly, in order to determine if range  $R_i$  qualifies as a small range, the time span of range  $R_i$  selected in step 2304 is compared to threshold time value "SmallRangeSize" (step 2306). The value of *SmallRangeSize* may be user-configurable. According to an embodiment of the present invention, *SmallRangeSize* is set to 8 seconds.

[254] If it is determined in step 2306 that the range  $R_i$  selected in step 2304 does not qualify as a small range (i.e., the time span ( $R_E - R_S$ ) of range  $R_i$  is greater than the threshold value *SmallRangeSize*), then the range is not a candidate for combination with another range. The value of variable "i" is then incremented by one (step 2308) to facilitate selection of the next range in the set of "N" ranges. Accordingly, according to the teachings of the present invention depicted in Fig. 23, only ranges that qualify as small ranges are eligible for combination with other neighboring ranges.

[255] After step 2308, server 104 determines if all the ranges in the set of "N" ranges have been processed. This is done by determining if the value of "i" is greater

than the value of "N" (step 2310). If the value of "i" is greater than "N", it indicates that all the ranges in the set of ranges for the multimedia information have been processed and processing of flowchart 2300 ends. If it is determined in step 2310 that "i" is less than or equal to "N", then it indicates that the set of "N" ranges comprises at least one range that has not been processed according to flowchart 2300. Processing then continues with step 2304 wherein the next range  $R_i$  is selected.

[256] If it is determined in step 2306 that range  $R_i$  selected in step 2304 qualifies as a small range (i.e., the time span ( $R_E - R_S$ ) of range  $R_i$  is less than or equal to the threshold value *SmallRangeSize*), the present invention then performs processing to identify a range that is a neighbor of range  $R_i$  (i.e., a range that occurs immediately before or after range  $R_i$  selected in step 2304) with which range  $R_i$  can be combined. In order to identify such a range, server 104 initializes variables to facilitate selection of ranges that are neighbors of range  $R_i$  selected in step 2304 (step 2312). A variable "j" is set to the value ( $i + 1$ ) and a variable "k" is set to the value " $(i - 1)$ ". A variable "j" is used to refer to a range that is a neighbor of range  $R_i$  and occurs after range  $R_i$ , and a variable "k" is used to refer to a range that is a neighbor of range  $R_i$  and occurs before range  $R_i$ . Fig. 24 depicts a simplified diagram showing the relationship between ranges  $R_i$ ,  $R_j$ , and  $R_k$ . As shown in Fig. 24, range  $R_i$  occurs after range  $R_k$  (i.e.,  $R_S$  of  $R_i$  occurs after  $R_E$  of  $R_k$ ) and before range  $R_j$  (i.e.,  $R_E$  of  $R_i$  occurs before  $R_S$  of  $R_j$ ).

[257] Server 104 then determines if the set of "N" ranges created for the multimedia information includes a range that is a neighbor of range  $R_i$  selected in step 2304 and occurs before range  $R_i$ , and a range that is a neighbor of range  $R_i$  and occurs after range  $R_i$ . This is done by determining the values of variables "j" and "k". If the value of "j" is greater than "N", it indicates that the range  $R_i$  selected in step 2304 is the last range in the set of "N" ranges created for the multimedia information implying that there is no range that occurs after range  $R_i$ . If the value of "k" is equal to zero, it indicates that the range  $R_i$  selected in step 2304 is the first range in the set of "N" ranges created for the multimedia information implying that there is no range that occurs before range  $R_i$ .

[258] Accordingly, server 104 determines if range  $R_i$  has a neighboring range that occurs before  $R_i$  and a neighboring range that occurs after  $R_i$ . This is done by determining if the value of "j" is less than "N" and if the value of "k" is not equal to zero (step 2314). If the condition in step 2314 is satisfied, then it indicates that the set of "N" ranges comprises a range that is a neighbor of range  $R_i$  selected in step 2304 and occurs before range  $R_i$ , and a range that is a neighbor of range  $R_i$  and occurs after range  $R_i$ . In this



case, processing continues with step 2316. If the condition in step 2314 is not satisfied, then it indicates that range  $R_i$  selected in step 2304 is either the first range in the set of "N" ranges implying that there is no range that occurs before range  $R_i$ , and/or that range  $R_i$  selected in step 2304 is the last range in the set of "N" ranges implying that there is no range that occurs after range  $R_i$ . In this case, processing continues with step 2330.

[259] If the condition in step 2314 is determined to be true, server 104 then determines time gaps between ranges  $R_i$  and  $R_k$  and between ranges  $R_i$  and  $R_j$  (step 2316). The time gap (denoted by  $G_{ik}$ ) between ranges  $R_i$  and  $R_k$  is calculated by determining the time between  $R_S$  of range  $R_i$  and  $R_E$  of  $R_k$ , (see Fig. 24) i.e.,

$$G_{ik} = (R_S \text{ of } R_i) - (R_E \text{ of } R_k)$$

The time gap (denoted by  $G_{ij}$ ) between ranges  $R_i$  and  $R_j$  is calculated by determining the time between  $R_E$  of range  $R_i$  and  $R_S$  of  $R_j$ , (see Fig. 24) i.e.,

$$G_{ij} = (R_S \text{ of } R_j) - (R_E \text{ of } R_i)$$

[260] According to the teachings of the present invention, a small range is combined with a neighboring range only if the gap between the small range and the neighboring range is less than or equal to a threshold gap value. The threshold gap value is user configurable. Accordingly, server 104 then determines the sizes of the time gaps to determine if range  $R_i$  can be combined with one of its neighboring ranges.

[261] Server 104 then determines which time gap is larger by comparing the values of time gap  $G_{ik}$  and time gap  $G_{ij}$  (step 2318). If it is determined in step 2318 that  $G_{ik}$  is greater than  $G_{ij}$ , it indicates that range  $R_i$  selected in step 2304 is closer to range  $R_j$  than to range  $R_k$ , and processing continues with step 2322. Alternatively, if it is determined in step 2318 that  $G_{ik}$  is not greater than  $G_{ij}$ , it indicates that the time gap between range  $R_i$  selected in step 2304 and range  $R_k$  is equal to or less than the time gap between ranges  $R_i$  and  $R_j$ . In this case processing continues with step 2320.

[262] If it is determined in step 2318 that  $G_{ik}$  is not greater than  $G_{ij}$ , server 104 then determines if the time gap ( $G_{ik}$ ) between range  $R_i$  and range  $R_k$  is less than or equal to a threshold gap value "*GapThreshold*" (step 2320). The value of *GapThreshold* is user configurable. According to an embodiment of the present invention, *GapThreshold* is set to 90 seconds. It should be apparent that various other values may also be used for *GapThreshold*.

[263] If it is determined in step 2320 that the time gap ( $G_{ik}$ ) between range  $R_i$  and range  $R_k$  is less than or equal to threshold gap value *GapThreshold* (i.e.,  $G_{ik} \leq$

*GapThreshold*), then ranges  $R_i$  and  $R_k$  are combined to form a single range (step 2324). The process of combining ranges  $R_i$  and  $R_k$  involves changing the end time of range  $R_k$  to the end time of range  $R_i$  (i.e.,  $R_E$  of  $R_k$  is set to  $R_E$  of  $R_i$ ) and deleting range  $R_i$ . Processing then continues with step 2308 wherein the value of variable "i" is incremented by one.

[264] If it is determined in step 2320 that time gap  $G_{ik}$  is greater than *GapThreshold* (i.e.,  $G_{ik} > \text{GapThreshold}$ ), it indicates that both ranges  $R_j$  and  $R_k$  are outside the threshold gap value and as a result range  $R_i$  cannot be combined with either range  $R_j$  or  $R_k$ . In this scenario, processing continues with step 2308 wherein the value of variable "i" is incremented by one.

[265] Referring back to step 2318, if it is determined that  $G_{ik}$  is greater than  $G_{ij}$ , server 104 then determines if the time gap ( $G_{ij}$ ) between ranges  $R_i$  and  $R_j$  is less than or equal to the threshold gap value "*GapThreshold*" (step 2322). As indicated above, the value of *GapThreshold* is user configurable. According to an embodiment of the present invention, *GapThreshold* is set to 90 seconds. It should be apparent that various other values may also be used for *GapThreshold*.

[266] If it is determined in step 2322 that the time gap ( $G_{ij}$ ) between ranges  $R_i$  and  $R_j$  is less than or equal to threshold gap value *GapThreshold* (i.e.,  $G_{ij} \leq \text{GapThreshold}$ ), then ranges  $R_i$  and  $R_j$  are combined to form a single range (step 2326). The process of combining ranges  $R_i$  and  $R_j$  involves changing the start time of range  $R_j$  to the start time of range  $R_i$  (i.e.,  $R_S$  of  $R_j$  is set to  $R_S$  of  $R_i$ ) and deleting range  $R_i$ . Processing then continues with step 2308 wherein the value of variable "i" is incremented by one.

[267] If it is determined in step 2322 that time gap  $G_{ij}$  is greater than *GapThreshold* (i.e.,  $G_{ij} > \text{GapThreshold}$ ), it indicates that both ranges  $R_j$  and  $R_k$  are outside the threshold gap value and as a result range  $R_i$  cannot be combined with either range  $R_j$  or  $R_k$ . In this scenario, processing continues with step 2308 wherein the value of variable "i" is incremented by one.

[268] If server 104 determines that the condition in step 2314 is not satisfied, server 104 then determines if the value of "k" is equal to zero (step 2330). If the value of "k" is equal to zero, it indicates that the range  $R_i$  selected in step 2304 is the first range in the set of "N" ranges created for the multimedia information which implies that there is no range in the set of "N" ranges that occurs before range  $R_i$ . In this scenario, server 104 then determines if the value of variable "j" is greater than "N" (step 2332). If the value of "j" is also greater than "N", it indicates that the range  $R_i$  selected in step 2304 is not only the first range but also

the last range in the set of "N" ranges created for the multimedia information which implies that there is no range in the set of ranges that comes after range  $R_i$ . If it is determined in step 2330 that "k" is equal to zero and that "j" > N in step 2332, it indicates that the set of ranges for the multimedia information comprises only one range (i.e.,  $N = 1$ ). Processing depicted in flowchart 2300 is then ended since no ranges can be combined.

[269] If it is determined in step 2330 that "k" is equal to zero and that "j" is not greater than "N" in step 2332, it indicates that the range  $R_i$  selected in step 2304 represents the first range in the set of "N" ranges created for the multimedia information, and that the set of ranges includes at least one range  $R_j$  that is a neighbor of range  $R_i$  and occurs after range  $R_i$ . In this case, the time gap  $G_{ij}$  between range  $R_i$  and range  $R_j$  is determined (step 2334). As indicated above, time gap  $G_{ij}$  is calculated by determining the time between  $R_E$  of range  $R_i$  and  $R_S$  of  $R_j$ , i.e.,

$$G_{ij} = (R_S \text{ of } R_j) - (R_E \text{ of } R_i)$$

Processing then continues with step 2322 as described above.

[270] If it is determined in step 2330 that "k" is not equal to zero, it indicates that the range  $R_i$  selected in step 2304 represents the last range in the set of "N" ranges created for the multimedia information, and that the set of ranges includes at least one range  $R_k$  that is a neighbor of range  $R_i$  and occurs before range  $R_i$ . In this case, the time gap  $G_{ik}$  between range  $R_i$  and range  $R_k$  is determined (step 2336). As indicated above, time gap  $G_{ik}$  is calculated by determining the time gap between  $R_S$  of range  $R_i$  and  $R_E$  of  $R_k$ , i.e.,

$$G_{ik} = (R_S \text{ of } R_i) - (R_E \text{ of } R_k)$$

Processing then continues with step 2320 as described above.

[271] Fig. 25A depicts a simplified diagram showing a range created by combining ranges  $R_i$  and  $R_k$  depicted in Fig. 24 according to an embodiment of the present invention. Fig. 25B depicts a simplified diagram showing a range created by combining ranges  $R_i$  and  $R_j$  depicted in Fig. 24 according to an embodiment of the present invention.

[272] As indicated above, the processing depicted in Fig. 23 may be performed after one or more ranges have been created according to the times associated with the hits according to flowchart 2200 depicted in Fig. 22. According to an embodiment of the present invention, after the ranges have been combined according to flowchart 2300 depicted in Fig. 23, the ranges may then be displayed to the user in GUI 2000 according to step 2108 in Fig. 21.

[273] According to an alternative embodiment of the present invention, after combining ranges according to flowchart 2300 depicted in Fig. 23, a buffer time is added to

the start time and end time of each range. A user may configure the amount of time (*BufferStart*) to be added to the start time of each range and the amount of time (*BufferEnd*) to be added to the end time of each range. The buffer times are added to a range so that a range does not start immediately on a first hit in the range and stop immediately at the last hit in the range. The buffer time provides a lead-in and a trailing-off for the information contained in the range and thus provides a better context for the range.

[274] A buffer is provided at the start of a range by changing the  $R_S$  time of the range as follows:

$$R_S \text{ of range} = (R_S \text{ of range before adding buffer}) - \text{BufferStart}$$

A buffer is provided at the end of a range by changing the  $R_E$  time of the range as follows:

$$R_E \text{ of range} = (R_E \text{ of range before adding buffer}) + \text{BufferEnd}$$

[275] Fig. 26 depicts a zoomed-in version of GUI 2000 depicting ranges that have been automatically created according to an embodiment of the present invention. A plurality of hits 2602 satisfying criteria provided by the user are marked in thumbnail 2008-1 that displays text information. According to an embodiment of the present invention, the hits represent words and/or phrases related to user-specified topics of interest. As depicted in Fig. 26, two ranges 2006-2 and 2006-3 have been automatically created based upon locations of the hits. Range 2006-2 has been created by merging several small ranges according to the teachings of the present invention (e.g., according to flowchart 2300 depicted in Fig. 23).

[276] Although specific embodiments of the invention have been described, various modifications, alterations, alternative constructions, and equivalents are also encompassed within the scope of the invention. The described invention is not restricted to operation within certain specific data processing environments, but is free to operate within a plurality of data processing environments. Additionally, although the present invention has been described using a particular series of transactions and steps, it should be apparent to those skilled in the art that the scope of the present invention is not limited to the described series of transactions and steps. For example, the processing for generating a GUI according to the teachings of the present invention may be performed by server 104, by client 102, by another computer, or by the various computer systems in association.

[277] Further, while the present invention has been described using a particular combination of hardware and software, it should be recognized that other combinations of hardware and software are also within the scope of the present invention. The present invention may be implemented only in hardware, or only in software, or using combinations thereof.

[278] The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense. It will, however, be evident that additions, subtractions, deletions, and other modifications and changes may be made thereunto without departing from the broader spirit and scope of the invention as set forth in the claims.

10081129 022402  
201222 6211801